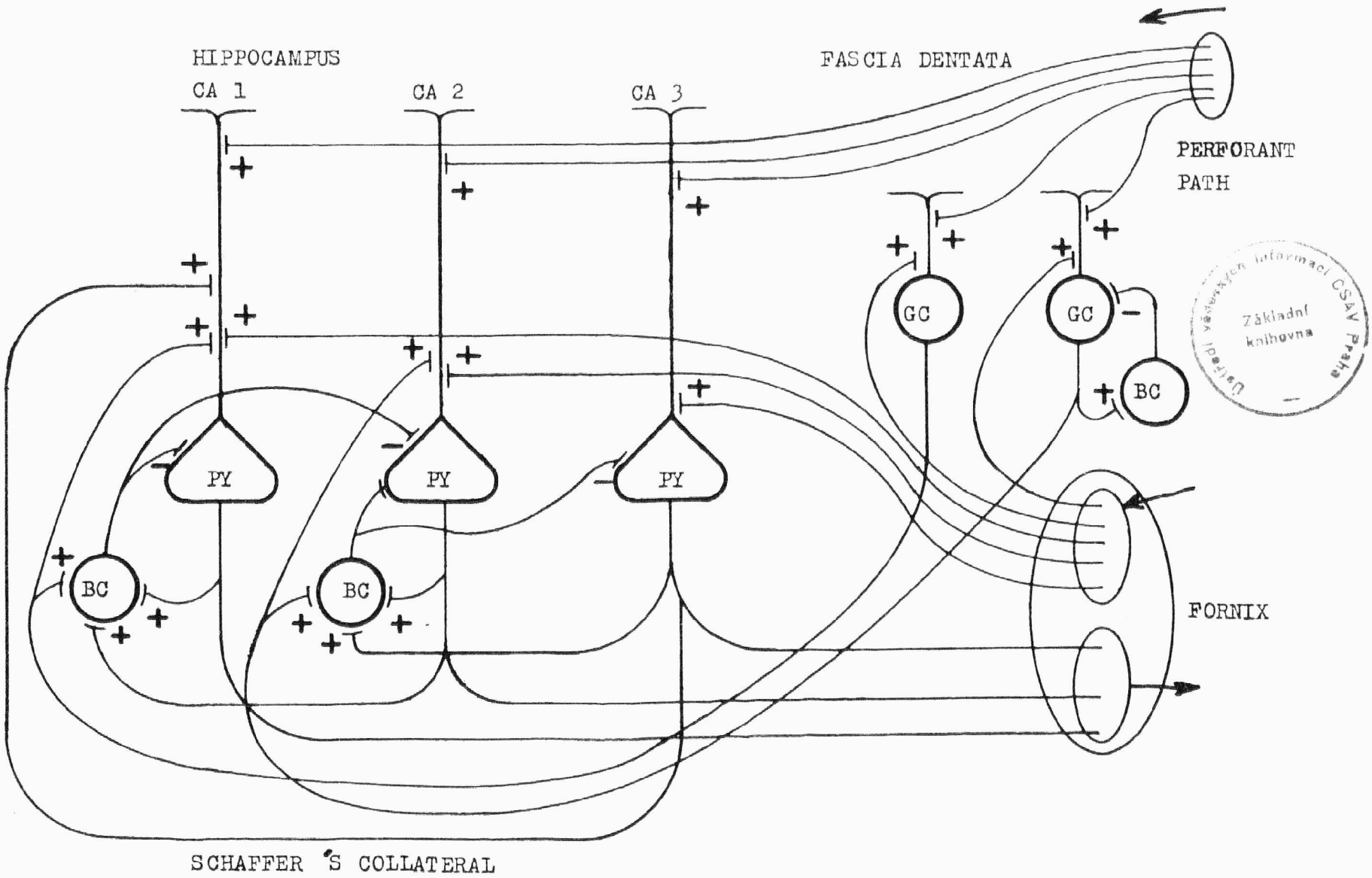
IDG

NEURAL NETWORK WORLD

International Journal on Neural and Mass-Parallel Computing and Information Systems

VOLUME 1 **1991** **NUMBER 1**



NEURAL NETWORK WORLD is published in 6 issues per annum by the Computer World Company, Czechoslovakia, 120 00 Prague, Blanická 16, Czechoslovakia, the member of the IDG Communications, USA.

Institute of Computer and Information Science, Czechoslovak Academy of Sciences, 182 07 Prague, Pod vodárenskou věží 2, Czechoslovakia.

Phone: (004422) 82 16 39, (00422) 815 20 80, (00422) 815 31 00

Fax: (00422) 85 85 789,

E-Mail: CVS35@CSPGCS11. BITNET

Price Information:
Subscription rate 399 US $ per annum.
One issue price: 66.50 US $.
Subscription adress: IDG Co., Czechoslovakia,
120 00 Prague 2, Blanická 16, Czechoslovakia.

Advertisement: Ms. M. Váňová, Ms. Ing. H. Vančurová,
IDG Co., Czechoslovakia, 120 00 Prague 16, Blanická 16
Phone: (00422) 25 80 23, Fax: (00422) 25 73 59.

Scanning the Issue:

ISSN 46 999

Editorial

THE INAUGURATION OF A NEW JOURNAL

We welcome you to the first issue of Neural Network World, the first scientific journal published in Middle and Eastern Europe devoted to the problems of neural and mass-parallel computing and information systems. The field of neurocomputing, which roughly speaking involves a good part of this area, is in recent years the one having the steepest increase of interest around the world. Though several well known journals dedicated to neural networks and neurocomputers already existed for a few years (and several more appeared last year), none of them comes from this part of the world, where a considerably large interest of many people in this area has a good history. Due to the political changes in our country and thanks the support of the IDG, Company, Czechoslovakia in Prague, we were able to prepare in very short time the foundation of the journal, in which we hope to create a scientific forum for the free exchange of ideas, knowledge and meanings of all the people interested in neuroscience and in related scientific areas. We hope that in addition to the people living in this part of the world our colleagues from other geographical areas will also contribute to Neural Network World.

The field of neuroscience is very wide. We can take it as dramatic and dynamic scene, illuminated by many beams emitted from different sources, in which the knowledge and experience of various areas comes together. There is a great variety of these sources and it is hard to identify all of them. However, definitely included among we find neurophysiology, mathematics, physics, informatics , cybernetics and electronics (see *Fig. 1*). Of course, besides these, there are other sources, such as microbiology, molecular genetics on one side, psychology and linguistics on the other side and chemistry, communication engineering and various engineering applications intersecting with the beams illuminating the scene, where the authors the neuroscientists -have to play much more complicated roles. The beams have different colors, as the languages in which the knowledge from individual sources is emitted are not identical. There are also not all of point nature. Some of them, like the source of mathematics e. g. , consist of several grouped partial sources. Some have a diffusion character. Therefore it is very difficult to define their intersection on the scene and to identify the actual color of the particular place, where one or the other actor is acting.

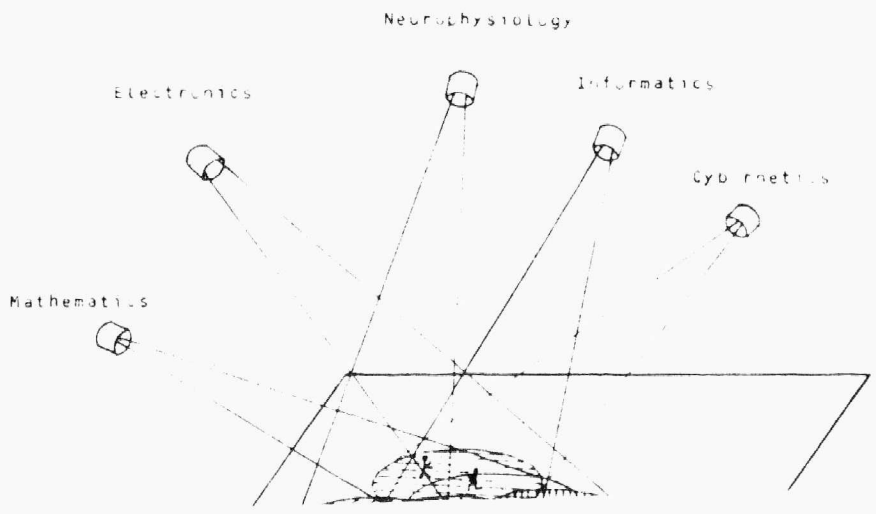


Fig. 1. The *ıl*

Moreover, the whole scene is not static it changes very dynamically. The sources emit their beams of knowledge in different parts of the scene, changes their intensity in time (sometime also stop), switch position (comes closer together or take distance) and also the actors do not play all their roles by staying in the same place in the scene. The whole scene of neuroscience can therefore be considered by an independent observer as a „Big Chaos". This probably happens often especially to the observers who are close to the bright and dazzling sources of knowledge illuminating the scene.

Nevertheless, there are positions from which at least some parts of the scene can be observed with satisfactory probability to recognize the sense of the play and to understand their internal relations and laws. I personally belong to the people who believe that the whole is not a large chaotic motion of the Brownian character, but that this a very dramatic and complicated development of the deep and old tendency of the human being to try to understand at least a little bit of its own consciousness and nature.

What role can be played here by the scientific journals specialized more or less to this field?

Last fall, after a considerably successful International Symposium on Neural Networks and Neurocomputing NEURONET '90 held in Prague and having been asked by Vladimir Tichý, the manager of the IDG Company, Czechoslovakia and to whom I am very grateful for this idea and support in their development, to prepare the publication of a new scientific journal devoted to this field, I needed to try, at least for myself, to answer this question. I came to the feeling that for such a dramatic scene like the one demonstrated schematically in *Fig. 1*, such journals can represent the screens or mirrors on which not only the interested observers but also the actors can see some parts of the drama in more detail and fixed time.

Therefore, by the use of such tools, the actors — the number of whom is now already high enough — and also the observers can take advantage of having more information about the activity in the distant parts of the scene, which are from their own standpoint not quite clearly visible (see *Fig. 2*). Of course, there are

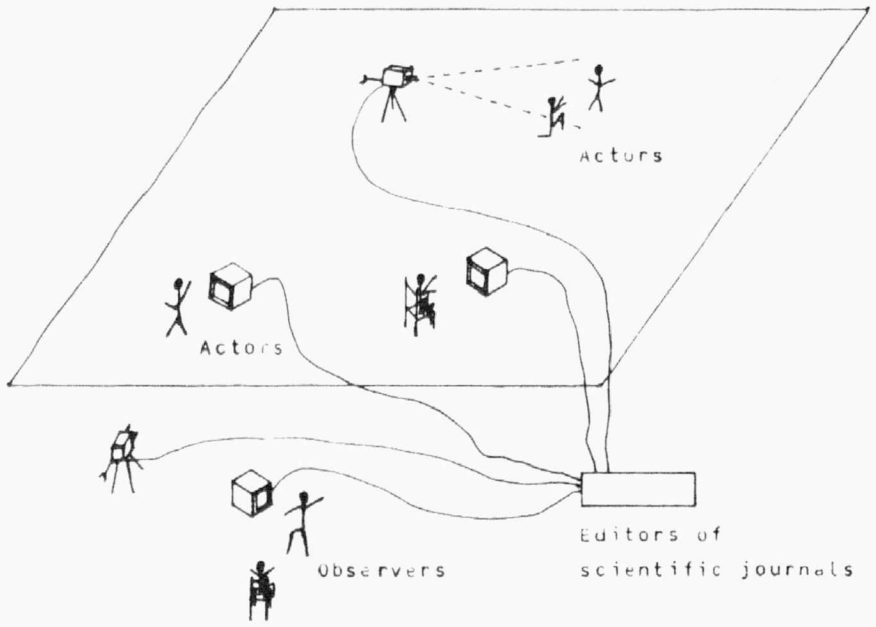


Fig. 2: To the role of scientific journals.

now at our disposal several kinds of such screens. Above all, we have here the group of well — known high level scientific journals. By the end of 1990, about 6 of them were known to me. These are:

Neural Networks, by Pergamon Press

International Journal of Neural Systems, by World Scientific,

IEEE Transactions on Neural Networks, by IEEE,

Neural Computing, by MIT Press,

Neural Technology Update, by Elsevier and

International Journal on Neural Networks.

Of course, this list is not complete and I am also sure that in 1991 more such journals will appear. Beside in these, the papers concerning neuroscience problems appear almost regularly in many other scientific journals, like e. g. in the

IEEE Transactions on Systems, Man and Cybernetics,

IEEE Transactions on Acoustics, Speech and Signal Processing,

IEEE Transactions on Pattern Analysis and Machine Intelligence

IEEE Transactions on Circuit and Systems,

IEEE Transactions on Information Theory,

IEEE Control System Magazine,

IEEE Circuit and Device Magazine,

IEEE ASSP Magazine,

Proceedings of the IEEE,

Computer,

Spectrum,

Journal on Cognitive Neuroscience,

Journal of New Generation Computer Systems,

Communications of the ACM,

Biological Cybernetics,

Expert Systems,

Simulation,

The American Physical Society,

Applied Intelligence,

Applied Optics,

Optical Letters,

Physica,

Nature, etc.

All of these periodicals have their own profile and contribute to the information of both the actors and observers of the neuroscience scene.

Therefore the second question which we need to answer is:

"Is there really space (and if there is, where?) for a new scientific journal in this area?"

The answer to this question is much more difficult. It can be influenced above all by the fact that at present, the total publication capacity of all these periodicals is still evidently much smaller than corresponds to the quickly increasing number of people active in it and having results worth publishing (even if one considers the necessary amount of rejected or long corrected papers). The present-day reality is that many authors have to wait many months for publication and that some results lose therefore a part of their timeliness. The second aspect influencing the answer comes from the regional distribution of the existing journals. Up to now all of them are published in the western hemisphere. However there are also other parts of the world, where the political barriers are fortunately now disapearing where there are many people who are interested in this field who can contribute to its development.

Being here in Prague in the heart of Europe just between these two approaching parts of the world, we would like to offer to all of these people the better possibility to publish their ideas and research results and we hope that they will make use of it. We hope also, that such a screening of activity in neuroscience in the East will be interesting for the people from the West and vice versa, so that our western colleagues will present in Neural Network World some of their results which should be relayed here quicker.

Summarizing these and some further aspects, we do hope that we can find our right place between all the other screens and mirrors on the scene.

The scope of Neural Network World.

Nevertheless, the scene of neuroscience is not a flat plane. It can be better compared to the relief of the landscape with many hills and vallies. There is almost impossible to mirror such a complicated object in one screen. Therefore one needs in any case to concentrate one's interest on some parts of the scene. The question is, how to choose them, how to place the cameras and how to display the pictures on the screen. Evidently, if the focus will be too narrow, the consciousness of the context can be lost.

We expect that the reasonable compromise between a too wide and a too narrow scope of our Journal will

be in focusing it on that aspect of neuroscience which concerns the processing, saving and transformation of information. Therefore we would like to have among our contributors especially those authors of papers dealing with the

theory of neural networks, natural and artificial,
methods of neurocomputing,
parallel computing methods,
synthesis and construction of neurocomputers,
distributed and parallel computer systems,
mass-parallel information processing,
biophysics and neuroscience,
applications of neurocomputing in science and engineering.

Of course, the selection of good papers is not an easy task and therefore we hope that the International Editorial Board of our Journal, membership in which was accepted by well – known people in neuroscience, will be a great help to us. We are grateful to many of them for their support and encouragement without which we would hardly be able to start this Journal in such considerably short time.

The structure of Neural Network World.

Like some other scientific journals, the profile of which we have known for many years, we would like to offer to our readers not only the presentation of interesting and significant papers, improving the general knowledge of neuroscience, but also some other useful services. Of course, we shall accept for publication in Neural Network World not only the regular **Papers** of the usual size 20 to 35 pages, but also the short contributions of a few pages' size, which we shall present as **Letters** (I know very well from my own long experience in circuit and systems theory, that quite often very important ideas appear just in such short messages).

Besides this, we shall insert in some of our issues a **Tutorial** section, in which we would like to present to the readers (especially to those who are just starting their interest in this field) some useful tools, like survey views, descriptions of algorithms and information on computer programs.

The almost standard service in any scientific journal consists in the presentation of a **Literature survey** and **Book reviews**. However, we think that especially in our field of neuroscience this is of special importance because of its high dynamics. We shall take a great care with it and try to inform the readers about all the interesting presentations which come in the Scientific Information System of the Institute of Computer and Information Science of the Academy of Sciences in Prague, with which we cooperate in this respect.

We also would like to inform the readers about the most interesting **Coming Events** in neuroscience, about the **New Products** appearing in the market and about the **Neurocomputer Companies**, the number of which is substantially increasing.

Such a structure of our Journal is of course not a rigid skeleton. It will be modified according to the experience with respect to the development of the whole field of neuroscience.

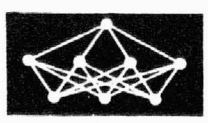
Appreciation

The creation of a new scientific journal in neuroscience is an exciting, but complicated and not easy activity.

The undertakings are large and I personally feel quite strongly the responsibility to do all the necessary things well. Therefore I am very indebted to all the people who have helped us, above all to the members of the International Editorial Board, to the contributors, to the staff of the IDG Company Czechoslovakia in Prague and to all my colleagues working as Associate Editors. They have done an excellent work without which we could not succeed to publish the first issue of Neural Network World in such tight time schedule.

We hope that the result of this hard work, which we now are giving into your hands, our respectful reader, will be interesting and useful for you and that you will appreciate our efforts.

Mirko Novák
Editor — in — Chief



CAN NEURAL NETWORKS EVER BE MADE TO THINK?

J. G. Taylor)*

Abstract:
An outline is given of neural network modules, and the modes of them, such that a machine operating with such a structure can be said to be thinking. The approach is based on a relational theory of meaning, in which the relations are determined by developing episodic memory in the net. This later form of memory is itself based on temporal sequences and their storage, as is the possibility of the machine developing "trains of thought".

1. Introduction

Thinking has so far been regarded in the main as an activity arising from processes occurring in living things and not in machines. This is in spite of the vigorous attacks on the problem of the nature of thought through artificial intelligence, especially its more recent developments in semantics, natural language processing, functionalism and connectionism. Thinking is thus still elusive; in this paper we propose to look at it from a new angle.

One can identify three major modes of thinking, concerned with: problem solving (directed), daydreaming (autistic) or memorising (mnemonic). It might be expected that if a machine can be built to function on one or other of these modes then it can be said to be thinking. AI programs which prove, say, theorems in geometry are thinking in a directed manner. Artificial neural nets, which can be trained effectively to classify inputs in various ways, and so be said to 'remember them', can be claimed to be thinking in a mnemonic mode. But each of these sorts of machines is performing in a very circumscribed manner. The symbol manipulation machine is not yet capable of powers of an ANN as far as associative memory is concerned, nor is an ANN capable of inferential processes like an AI machine. There are various recent attempts to form hybrids of neural nets with rule-based systems so as to extract the benefits of each. Yet even these hybrids do not seem to have autistic powers. The purpose of this article is to explore the possibility of going beyond the separate and rather low level of activities of present-day neural net classifiers and AI machines. We wish to consider the possibility of a machine which would be said to be thinking at a higher level since it involves if possible all three modes of thought simultaneously. The thesis to be argued here is that there is a general approach to the construction of a thinking machine which seems to allow an analysis of thinking at a level comparable to that of humans. In particular it leads to the notion of "meaning" of states and inputs which gives a new slant to that concept from a more general point of view.

The problem of meaning of inputs is seen as a crucial one for the AI approach to the modelling of thought. Trenchant criticisms of the lack of semantics, as compared to syntax, have been made by various authors, especially Searle [1] in his 'Chinese room' moral, and by Patricia Churchland [2]. These criticisms do seem correct, for semantics can never be given purely by the fixed set of rules obeyed by predicates in a logical calculus. There is more to meaning than can be determined by logical manipulation. Yet developments have occurred in both logical and AI semantics in which the meaning of propositions is beginning to be included. The conditions (in model conditional semantics) or the situations (in situational semantics) in which the validity of the proposition is to be assessed are currently being incorporated [3]. It is fair to say that these approaches still require a great deal of development before they can be said to begin to model human thought. It is of value to note that there turns out to be considerable similarity between these developments and our approach, which arises from the modelling of neural nets.

One should also take seriously the empirical observation that the AI program has not been as effective as had originally been hoped; for example the move from the fifth to the sixth generation computer by the Japanese.

Other approaches to thinking have been tried which are related to AI. Functionalism is presently attracting interest. The proposition that mental states are satisfactorily characterised by their causal role is illuminating. However it does not seem to pass the acid test of leading to a blueprint for constructing a thinking machine. For only if one can do that, or indicate how it might be achieved in principle, may one's theory begun to be put to any experimental test. It is not at all clear to me how to draw a blueprint for a machine operating on functional lines to lead to activity which could be termed 'thinking' at a non-trivial level. To appeal to the fact that the brain acts in such a manner

*) Prof. J. G. Taylor
Department of Mathematics
King's College
Strand, London, WC2R2LS, U.K.

is to miss the point that a functional model may not be sufficient to allow for the construction of a thinking machine, and must therefore only be a very simplified description.

Similar structures apply to the strict Connectionist approach. This uses a semantic network, with nodes which are concepts and the strengths of connections between them determines the relations between them. Such a framework does not allow for any internal structure of the concepts which is necessary in order for deductive processing to be possible, as pointed out by Fodor and Pylyshyn [4] and by Clark [5]. The lack of any inferential structure thus leads to a system which is too weak to be able to describe human thought.

Neural networks seem, at first glance, to have a better chance of succeeding in developing a theory of meaning and even of human thought. Neural nets are claimed to be simplified models of the brain. At the same time they are much broader than the connectionist approach involving only semantic nets. However the simplicity may be a snare and a delusion: one should remember Einstein's dictum "one can make a model simple but not too simple". Present progress in benchmarking neural nets against traditional approaches has shown [6] that neural nets have only a slight edge over these latter methods. That is one reason why there has been a development towards taking a closer look at the complexity of the brain and trying to extract further algorithmic features from it than the purely logical or static features chosen so far. This is the program of reverse engineering. For example, stochastic and temporal features of neurons are presently being investigated [7] for their help in improving the processing power of artificial nets. The details of synaptic action [8] (in particular in terms of stochastic aspects [9]), of temporal features [10] and of structural aspects of single neurons [11] are all under the microscope in this respect.

These attempts to complexify neural networks, so as to have neural models closer to living brains, do not address the issue as to what are the overall programs being used by the brain. This may be ascribed to the fact that we just do not know what algorithms are being used by the various areas of the living cortex. Thus even the nature of the processing is uncertain in the different layers of areas 17 and 18 of striate cortex in the cat or monkey, in spite of the effort spent into research on that question [12]. The recently discovered cortical oscillations [13] give a hint as to the formation and dissolution of temporally synchronised neuronal cell assemblies as evoked by suitable external stimuli. But the detailed learning processes for suitable connectivities and the nature of further information processing in associative cortex is still unknown. Similar ignorance exists about the olfactory cortex.

Such a level of ignorance does not mean that we should adopt the position of the 'boggled sceptic' of Patricia Churchland [2]. Artificial neural net analysis involves a general attempt to analyse information processing by neurons at all levels. Thus it is indeed appropriate to attempt to understand how meaning, or other features of human thinking, may in principle be implemented or arise out of neural net activity of a broad range of forms. That is the basic problem considered in this paper. We will use the presently understood neural net properties to try to discover the architectures and modes of action which could lead to "thinking" neural nets. At the same time we will attempt to link up with the strong AI approach to thought.

How can meaning be incorporated in the activity of a machine? The problem we are faced with is to determine the minimum structure which must be added to the simplest notion of machine in order for it to be able to begin to assign meaning to some of its inputs. Such meaning will be taken here to be determined by means of the following general 'overlap' thesis:

The meaning of an input is to be determined by the degree of overlap such an input has with previous inputs, when appropriately stored.

This thesis appears to be both reasonable and minimal. It is reasonable in the sense that any experience of an organism which only involves components of the input which have never been experienced by that organism before may be expected to have no or little meaning to that organism. The latter will be at a loss as to how to respond to such novelty; the phrase "lost its bearings" would be appropriate to describe the organism at that instant.

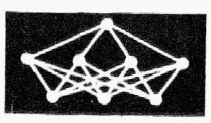
The thesis is minimal in that only comparison of incoming input with earlier inputs is required, say in terms of some distance between states. It is difficult to conceive of any simpler measure of meaning which can be quantified. Of course the overlap thesis has to be fleshed out in some detail in order for its implementations and implications to be appreciated. That will be attempted in this paper, first at a rather abstract level for automata in the next section, and then in more detail for neural network implementation in succeeding sections.

2. Machines with Meaning

A machine is taken here to mean an automaton $< I, O, S, \lambda, \delta >$, where $I(O)$ is the input (output) set, S is the state space and λ, δ are the next state and output functions

$$\lambda : I \times S \rightarrow S$$
$$\delta : I \times S \rightarrow O$$

with $\lambda(i, s)$ being the next state when the input is received by the machine when in state s, and $\delta(i, s)$ is the resulting output. Such a framework contains neural network dynamics, although the automata in question may need to function in continuous time and have an infinity of states and outputs in the realistic



case of graded neurons. Our discussion in this section will not use any finiteness criterion on any of the sets I, O, S.

In order to be able to implement the overlap thesis of meaning of the previous section we have to augment the automaton $A = <I, O, S, \lambda, \delta>$ by what we denote a memory function R. This must describe how a given state of A can evoke other states from memory which are relevant to 'coloring' the experience of that state. Thus R will be a function from S to the set of subsets of S (which we denote by 2^S):

$$R : S \to 2^S.$$

For a given state $s \in S$ the set of state in $R(s)$ will thus be the memories evoked by s. It is more general to take R also to depend on the inputs i, and that can be included if is so desired; we will not do that explicitly here. The automaton $<I, O, S, \lambda, \delta, R>$ will be termed a memorising machine. Some simple properties of R will be discussed later.

The overlap thesis can now be expressed in terms of R as follows. The meaning that A assigns to an input i and state s', which send the machine into the state $s = \lambda(i, s')$, is determined by the 'size' of $R(s)$. This size is an undefined concept, although it can be quantified in terms of a distance function on S; in that case the size of $R(s)$ would be its diameter. If S is finite then the size of $R(s)$ could be the number of elements it contains. In general if the size of $R(s)$ is denoted by $N(R(s))$ then the overlap thesis is:

The meaning that A assigns to s is determined by $N(R(s))$. (1)

In particular if $N(R(s)) = 0$ (zero) when $R(s) = 0$ then s can be said to be meaningless, as in the case remarked on in the previous section.

The overlap thesis can now be extended to more complex situations. In particular a very important question is that of the relative meaning of two succeeding inputs s_1, s_2. Suppose each of s_1, s_2 are meaningful, what is the degree of relative sense that the machine can make of them? That is relevant, for example, to the case when s_1 might be "this piece of cake" and s_2 is "is ill". Then the memories $R(s_1)$ would involve activities associated with eating, tasting, baking, etc and sensations described by various levels of gustatory enjoyment experienced while eating. The memories of $R(s_2)$ would involve those of places such as hospitals, beds, operating theatres and of people such as doctors and nurses. In this case there is no sense that can be made of the total sentence $s_1 s_2$, this being due to the fact that

$$R(s_1 s_2) = R(s_1) \cap R(s_2) = 0 . \quad (2)$$

The condition that $R(s_1 s_2)$ is null has been noted above as corresponding to $s_1 s_2$ having no sense to A, so the result (2) is consistent with earlier aspects.

More generally we can say that any two succeeding states (or inputs, on identifying states and inputs) of A make the degree of sense determined by $R(s_1) \cap R(s_2)$. Thus by (1) the amount of sense between two states is determined by the degree of overlap of their meanings. This agree with what is to be expected on the grounds of common sense.

No details have been given so far as to how R is to be specified; that will be done in the next section in the case of neural net implementation of A. Nor has there been any indication of how R will have developed so as to properly be regarded as a memory incorporating past experiences. There will have to be adaptive elements in S which are modified as experience occurs. Analysis of such plasticity is natural in the neural net case. However a minimal requirement for R is that if a state s is experienced at a time t_1 and also at a later time t_2 then, if no forgetting occurs,

$$R(s, t_1) = R(s, t_2) \quad (3)$$

where $R(s, t)$ denotes the memory function R at time t when in the state s. The strictly increasing relation in (3) is that for an efficient memory, and assuming A has experienced states other than s between t_1 and t_2. The equality relation is for a memory which has not been modified by more recent experiences between t_1 and t_2, such as might occur in the case of a patient with Korsakoff's syndrome. In such cases there could also be states for which $R(s, t)$ is a decreasing function of t, those being for which amnesia has occurred, when

$$R(s, t) = 0, \ t > t_0$$

for amnesic states s, with t_0 being the critical time of onset of the disease.

It is possible to extend the above machine to one with "mentation$_m$" (the subscript denoting only the mnemonic form) if λ and δ are taken as involving both s and $R(s)$:

$$\lambda, \delta : I \times S \to S, O . \quad (4)$$

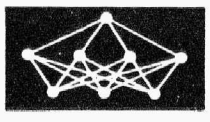
For then the 'mental content' of A when in the state s may be defined as:

$$M(s) = s \cup R(s) .$$

The set $M(s)$ will be an important subset for A if λ and δ are defined only in terms of $M(s)$, so

$$\lambda(M(s)) \in S, \delta(M(s)) \in O. \quad (5)$$

Such an epithet "mental" for $M(s)$ would seem to require strong justification after what was said in the introduction. That will only be attempted here as follows: the mental state of a human at any time is at least quite strongly determined by the amount of memory evoked by the state she/he is in at that time.



Thus when I look at a chair I see it 'colored' by the uses I have put it to in the past, by the remembrance of the feel of the surface texture, etc. If I see an object which I have never seen before I will have little mental content of that object; such content would be expected to relate directly to the degree of meaning the object (or the state it evokes in me) has for me.

We may, following this approach, define the level of consciousness at a time t for the machine A as

$$C(t) = N(M(s, t)). \tag{6}$$

It is then to be expected that the mental content C, averaged over a day, develops in time as in figure 1.

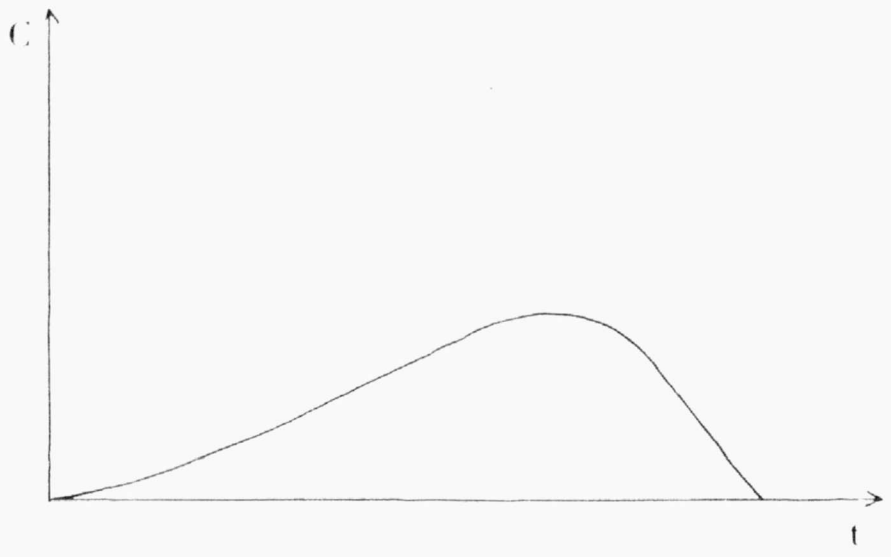


Fig. 1. The curve of dependence of the level of consciousness C(t) on the time for a typical individual. The slow early increase compared to the later sharper fall is due to the slower gathering of experience, compared to the more rapids decrease of consciousness level due to neural ageing and loss.

3. Meaning for Neural Networks.

A neural net is a set of N nodes which emit activities sent to the other neurons of the net (including themselves). These activities, which are denoted by $u_i(t)$ for that emitted by the i'th node at the time t, are accepted by the other nodes and contribute to their future activities. If discrete time steps are used then the general dynamical equations for the net will be

$$u_i(t + 1) = f_i(u(t), u(t - 1), \ldots, I_i(t)) \tag{7}$$

where f_i in general depends on the label i, $u(t)$ is the N-vector of activities of the net at time t and $I_i(t)$ is the input onto the i'th neuron at that time. Time delays have been included in the f_i, as have non-linear (Σ-Π) units. The connection weights in the general structure (7) are proportional to the partial derivatives of the f_i's with respect to their variables. The simplest case is when

$$f_i(u(t), \ldots) = f_i(\sum_{j,r} a_{ijr} u_j(t - r) + I_i(t)) \tag{8}$$

where f_i is a sigmoid type of function; the a_{ijr} are the connection weights for the time delay r. An even simpler case is when

$$a_{ijr} = a_{ij}(r = 0) = 0\,(r \neq 0) \tag{9}$$

when the net reduces to that standard in much of ANN modelling. The case (8) is important for our future discussions since it allows for direct temporal sequence storage (TSS) [10], which we will use as an important property of thinking nets.

ANN learning algorithms are based on three levels of supervision: none(unsupervised), rewarded by a global signal sent to all the neurons for certain of their outputs (Reinforcement training), or totally directed (supervised).

The goal of the latter form of learning is usually to reduce an error level, this being the difference between the desired and actual output. The reinforcement training aims to produce an output which maximises the reward given by the environment to the machine. Only the unsupervised learning scheme seems appropriate to describe the development of the initial stages of perceptual processing in humans, as catalogued, for example, in [14]. Higher levels of supervision undoubtedly then begin to play a role in the learning process, although the learning rules which vertebrate living systems use at a cortical level are still being vigorously explored.

The initial problem we face in considering neural net implementation of the memorising machine of the previous section is that of determining the general form the map R is expected to take. To do that we will build on the many discussions of the systems approach to the brain to take as the basic model an initial semantic network W (to be explained shortly) into which the input is fed, after some preprocessing at retinal and striate cortical level. This then feeds activity into an "episodic" memory net E (also to be explained shortly), as shown in figure 2. As in the previous section we will conflate states and inputs for W, so that states $w \in W$ may cause or bring about states $e(w) \in E$. Feedback could occur from E to W, but is not considered here.

The net W may be regarded as composed of a set of modules. Each of these gives an analysis of its input in a semantic manner. Thus the modules close to the preprocessed input build up low level categories from the input, whilst those which are further remote combine the latter into higher level ones comprising object categories. Such categories would be, say, of phonemes, syllables or words as auditory signals, of letters and words as visual signals, and so on. These concept modules would be heavily interlinked with each other by feed-forward and feed-back lines. A simple (probably too simple) method to build such semantic modules might be by means of Kohonen's topographic map [15], although this would have to be extended to the case of sets of nets interacting with each other as the learning proceeds. Such nets have not yet been built, but there seems no reason in principle why they cannot be so some decades in the future, when neural net technology has advanced that far.

The episodic net E develops at the same time as W. The evidence from the work of Penfield and others [16] is that episodic memory is stored as temporal sequences. This is made feasible by means of the hippocampus, following [17]. The method used there trains a net to store, and so be able to generate, a sequence of patterns by requiring the net output is a certain pattern of the sequence given that its input is the succeeding pattern. It was suggested in [17] how the architecture of the hippocampus can be seen as suitable for achieving such learning, and for subsequently generating a stored pattern sequence which could then be stored more permanently in the infero-temporal cortex. With about 5.10^7 neurons involved in hippocampus and new patterns arriving every 300 msecs it was estimated in [17] that the storage capacity would allow storage of pattern sequences for several weeks, a reasonable length of time. The involvement of hippocampus is not shown explicitly in *fig. 2* for reasons of clarity.

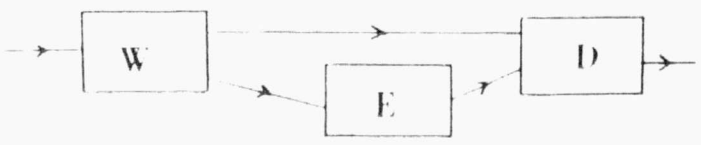


Fig. 2. Schematic of the connectivity of the machine with meaning. The net W is the semantic store, whilst E is the episodic net. The final net D is the decision unit, as discussed in the text.

There must be a close correlation between the nets E and W during learning. This allows the semantic analysis performed by W on inputs to be used both during the laying down of new memories and their reactivation in E. Such processes have not been achieved within ANN simulation, but yet again there apears to be no reason in principle why such learning cannot be achieved when the technology has advanced.

It is now necessary to determine a natural form which the mapping R of the previous section might take. To do that we will use the main feature of E already alluded to above, that an input w may 'activate' a set of possible memories $e(w)$ which are relevant to that input. By activate is meant one of several possibilities. The most obvious is that the different but relevant memories to w (relevant because they involve suitably large sets of objects initially coded by w and common to the input w) are indeed activated in E by W. Such activation may require a modular construction of E somewhat similar to W so that numbers of activities could be occurring simultaneously. On the other hand it could involve a competitive process in a single net, where activation of a number of different past memories can occur simultaneously in a distributed manner across the net. On the other extreme, the activation of these relevant memories in E may not be simultaneous, but could occur in a serial manner by some mechanism. However such a linear search would seem to slow down the use of the net E, so that such a mechanism does not seem to have good survival value.

It must be added here that the simplest possible model has been chosen for memory. Thus effectively there is a short term store W and a long term one, E. The distinctions between, for example, implicit and explicit memory are not being made here, but are features to be considered on more detailed analysis of our present model.

To return to R, we now consider how activation, of whatever form, of an episodic memory in E by a state w of W might be used by the system. One way to achieve that is to let the outputs of W and E (activated by w) be fed into some comparator or decision net D, also shown in figure 2. The decision net D may act, for example, in a competitive manner, so that it assesses the closeness of activated memories to their activating inputs, and gives an output determined not only by w but also by memories close enough to w. For it is these latter which should be relevant to the determination of responses to the input, as well as the satisfaction of goals (which have been left out of direct consideration here, again for simplicity).

It is now possible to define R in terms of the outputs w and $e(w)$ of W and E respectively to be of the form:

$$R(w) = \{e(w) : e(w) : = w\} \qquad (10)$$

where $: =$ denotes close to with respect to some distance function on the outputs of the nets W and E, regarded as inputs into a set of neurons in the decision net D. The relevant distance function may naturally be in terms of the activity arriving on the neurons of D, since that activity is what determines the manner in which D notices W and E. Thus if $A_i(e, w)$ is the activity of the i^{th} neuron of D due to the inputs w and e from the nets W and E respectively, then $e: = w$ may be taken as the condition

$$\max_i | A_i(e, 0) - A_i(0, w) | < \varepsilon \qquad (11)$$

for some suitably chosen ε. Since the response of the neuron is assumed to depend only on its activity, then the inputs e and w will be indistinguishable if ε is suitably small.

R defined as above corresponds to the condition that e and w look effectively the same to D. If e and w are too different then it is not expected that e should be involved in the mental state brought about by w. That is exactly what is achieved by the above construction. It is also possible to extend the above definition by allowing a range of values of ε to be used in (11), these values being under the control of other nets (by change of thresholds of the neurons in the net D). The construction of the previous section may now be used to define the 'mental' state of the system, the degree of meaning any input has to W, and the degree of relative meaning between two succeeding inputs (states) w_1, w_2. The net of figure 2 is thus a memorising machine and has the power to mentate$_m$. But it does not yet have the power to think.

4. Towards Thinking Neural Networks.

It is now necessary to consider in what manner we might be able to extend the construction of a machine with memory of section 2, implemented in neural net technology as in section 3, to that of a thinking machine. To achieve that end we must, according to the discussion in the introduction, incorporate the abilities of deductive and autistic thought. We will consider the former of these powers in this section and the latter in the following one.

The problem we are faced with here is to implement the "strong AI" approach to human thinking in neural net form. In effect we must understand how neural net architecture could be constructed so as to directly implement a suitable set of programmes of AI form. It is not that we wish to form nets of hybrid form, which partly incorporate distributed knowledge, as in the neural net dynamics of equation (8) and otherwise acts as an inference machine by running in terms of a programme in a suitable language using the outputs of the net as variables. What is being proposed here is much more drastic: we wish to construct nets which can also implement such AI programmes as a natural part of their activity qua neural networks, not qua serial computing machines. This can be properly called the "strong neural networks" approach to the problem of human thought.

Let us start by considering the essential elements of a formal language or an AI programme. This involves a set of symbols (the set of terminals in a formal language), a set of axioms for these symbols and the rules for the combination of the symbols to generate further collections of symbols which are to be regarded as valid theorems for the symbols (the productions or the derivation rules). Thus we might consider the symbols to be the integers, the axioms to be the Peano axioms and the productions as the associative, distributive and commutative rules of arithmetic.

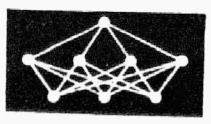
One of the important structural features of a formal language, say if it is regular or context-free, is that it may be defined as a non-deterministic automaton (NDA) or a non-deterministic pushdown automaton (NPDA) respectively. Thus we may be able to regard our problem as solved if we can implement either an NDA or an NPDA by means of the activity of a neural net. More generally we should consider how to implement a general Turing machine by a neural network. Let us consider solely here an NDA for regular grammars.

The finite-state non-deterministic automaton that can accept a regular language $L(G)$ of a regular phrase-structure grammar (N, T, P, S) is specified as follows. The system (N, T, P, S) has non-terminals N, terminals T, productions P and starting state S. The corresponding NDA has state space $NU\{*\}$ (where $\{*\}$ are the halting or final states), input alphabet T, state transition function t determined by the productions P in that, for $A \in N$, $a \in T$ requires that $t(A, a)$ contains *, and if $A \text{-->} aB$ is in P then $t(A, a)$ contains B. Thus the differences between the automata $<I, O, S, \lambda, \delta>$ considered earlier and the NDA $(NU\{*\}, T, t, S,\{*\})$ is the neglect of δ and O and the emphasis on the existence of the start state S and the final states *.

It is well known that any automaton of the form $<I, O, S, \lambda, \delta>$ may be explicitly implemented by means of a neural net. Can the same be achieved for NDAs of the above form? There seems to be no difficulty with the start state S; the following net activity from that initial state is to be regarded as accepting the sentences of the regular language $L(G)$. Activity derived from other initial states is to be disregarded. However the problem of ensuring that activity proceed solely to the final states does not appear trivial. Nor is the actual generation of required strings. It may be that these questions can be resolved in a natural manner, for example by training the net by a suitable algorithm to learn to accept only the allowed strings of the grammar, and so allow deductive or linguistic thought be implemented by neural nets. There is indeed a growing literature on the training of neural nets to implement NDA's [18]. We add parenthetically that it has not yet proved possible to train a net to implement a non-regular language, due to the arbitrary length of the associated stack automaton. It is clear that a neural net could never achieve such storage due to its finite capacity. However there cannot exist in any hardware device an infinite memory capacity, whether it be a neural net or a stack automaton. This leads one to suspect that in the human case we can only recognise or generate sentences of a bounded length. Indeed one of the essentials of good writing is to avoid longwinded sentences. So we should not regard the problem of infinite length stack memories as insuperable but only as one to make us strive for nets or other clases of memories with as great a capacity as possible.

Let us return to the question of the powers of neural nets trained as language automata, whether or not they also model stack memories of a limited length of pushdown automata. The main problem here is that such a net will only function as a recogniser; it will not be able to "think for itself". Moreover it seems necessary, in any case, to amplify the NDA so as to allow for the concept of meaning to be attached to the strings so generated, along the lines of the previous two sections. It would be preferred that language generation and logical thought be closely involved with the appropriate state of the semantic net W, since that has coded states intrinsically bound with linguistic or logical thoughts. In other words we need to look deeper for a construction of directed thought than by the rather naive attempt to make the net ape the action of an automaton implementing a grammar of a certain kind.

The states of the modules of the semantic net W should already contain the symbols of such thought, coded by suitable inter- and intra-connections. The basic feature to be added to the set of modules of W is thus the manner in which productions are achieved by



suitable connectivity. There seems to be two levels at which we can try to achieve this. The first is purely linguistic: how is language produced correctly by the nets? The key to that should be in the connectivity between various syllabic, phonemic and word modules in W. It is assumed that these are developed as, say, coupled topographic maps, so that excitation of only grammatically correct sentences is achieved with high probability for verbal production. Such correctness is expected to arise by training during the storing of sequences of words, in an unsupervised manner, in early childhood (2—6 years) and then later in a reinforced manner.

The second is at the level of the method of coding used by the episodic memory net, and the manner in which this is coupled to the semantic net W so as to give meaning to the words as they are being uttered. It is well-documented that the loss of Wernicke's area leads to deficit in meaning either in produced or received speech, and that Wernicke's area is connected to the temporal lobe, in which the episodic memory is thought to be stored. Thus we expect that the manner in which the episodic and semantic nets constrain each other will be crucial. Indeed it is known that loss of temporal lobe prevents the learning of new semantic material as well as new episodes.

So far there is no algorithm available which could be expected to allow the training of a neural net so that it can function at the level of a child as described above. There are two aspects to the problem of developing such training. Firstly it appears necessary to comprehend how temporal sequences, suitably preprocessed, could be stored effectively. This is expected to require the use of the hippocampus, since its loss is well documented to lead to lack of laying down of new experiences. But, as noted earlier, under such a circumstance there is a concomitant loss of new semantic material. This may be considered as involving the laying down of temporal sequences, but these may only be short. They will also be stored in the semantic regions W, so that there must be suitable interconnections from the hippocampus to W so as to cause an expansion of the various semantics nets mentioned earlier.

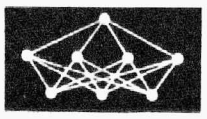
We have considered the way in which the hippocampus may store temporal sequences internally elsewhere [17]. It is now necessary to consider in what manner that organ may be used in guiding the storage of the two sorts of memory, semantic and episodic, in neocortex. We can only give a general outline here of a possible modus operandi, and leave to more detailed analysis, by simulation, the level of success that our proposal leads to.

It is to be expected that the general principles on which there is storage in both the semantic and episodic modules is the same. This follows from the rather similar architecture that neocortex possesses over its surface, outside the primary cortical regions. We thus assume that there is some level of competitive activity between local regions, as is evidenced from distributions of inhibitory collaterals in cortex [12]. At the same time we take account of the continued persistence of activity on the cell membrane for an important length of time, as evinced by the latest values of the membrane time constants for pyramidal cells [19]. We thus propose a temporal extension of the Kohonen topographic map [15]. In this the activity on the cell surface at a given time is used to determine the local winner, and the resulting updating of the weights. Since this activity involves earlier inputs, the topographic map which results will have built into it the temporal structure of the inputs. It is important to note that this extension of the Kohonen learning algorithm also involves the learning of lateral excitatory connections between the neurons of a module. It is these latter which will be of great importance in encoding the temporal structure of the input.

Such learning by means of temporal topographic maps will lead to strongly connected regions if they are strongly causally related. It appears that this connectivity will be an encoding of the rules of the grammar in the case of word and sentence modules.

It might be asked at this point as to the role of the hippocampus in the new learning processes described above. There appear to be at least two aspects which call for its presence. The first is to act as a medium term store for coded inputs from various modalities. The structure of the hippocampus allows it to function in this way more efficiently than neocortex, as is argued in [17]. The second is to mediate between underlying goals of the system, as stored in the hypothalamus, and the input as entering the hippocampus. There are good connections between these two organs, and it is one of the proposed functions of the interaction of the two that the importance of inputs for attaining emotional goals be mediated by the hippocampus. Thus the hippocampal feedback to the neocortex can be considered as involving a reinforcing activity for emphasising the laying down of permanent memories of the episodic (in temporal lobe) or semantic (in associative cortex) type.

We now claim that in the above manner the nets of the system will, if allowed to experience a suitable set of inputs, such as grammatically correctly produced sequences of words, develop a store of these words and sentences that encodes the grammar by means of the lateral connections both inside each net and between the modules. Such a claim does not appear to be contentious, in that the system is expected ultimately to achieve such storage. What is at issue is the training time and the storage capacity of the system. Indeed the former of these questions is still one of the main unsolved ones facing neural networks. That the system will scale correctly on increasing size of the input set is also a question facing natural language processing in AI, mentioned earlier. This question is not one we can expect to solve here, but will return to it elsewhere when we present the results of our simulations [20]. We add that a modular approach to speech



recognition allows for better scaling than if monolithic nets were to be used [21].

As to capacity, we have already discussed this for our hippocampal model in detail elsewhere [10], [22], and only note here that a storage capacity of the order of the number of neurons of the net, as shown in [10], [22], will lead to vocabularies of the order of hundreds of thousands of words, for nets of the size of those in the human brain. There may be gross degradation of the capacity in the case, for example, of patients with Korsakoff's or Alzheimer's disease; it could be of value to attempt to correlate in a quantitative manner what is known about the extent of cortical loss in these cases with memory degradation. Indeed these cases are of importance in that they may present cortices at the end of their capacity.

Our earlier claim that our system will have grammatical powers is now extended to the claim that it will have deductive powers. Given an input sequence, this will lead to activity in the semantic nets, controlled by the episodic net by means of the causal connections that have been built up in the learning process (the connections having been assumed to be present in the first place. The efficacy of such connections depends on the connection probability between appropriate pairs of neurons; that will be discussed elsewhere). Given that this activity can be output from the system in the manner described in section 3 (so by some sort of decision unit) and so as to return as later input, causal processing by the system on its own activity can proceed. We conjecture that this feedback-sequential activity may be used by the system in a manner which would correspond to deductive thought. The sentences will be produced whose content will be based on the earlier experience of the system. Moreover the system will be able to assign meaning to the inputs (and so to its own outputs) by the discussion of the earlier sections. It can thus be said to think in a manner in which the sentences it is using have meaning to them, which is used in the further development of its activity. It can be said to have meaningful deductive thought.

5. Creative Thought.

It is finally necessary to understand how it is possible to allow the system we have developed so far to think in an autistic or creative manner. By this we mean that the activities of the system are more controlled by its internal activity than from the exterior. It should be able to day-dream, so that its internal activities become the centre of its thought processes. Such a change of mode of activity has already been considered from the point of view of feedback control in a strictly hierarchical system in [23], and further developed in [24]. The model presented here may be extended along somewhat similar lines to those references so as to be in keeping with what is known about the general connectivity in the brain. However we should point out here the considerable difference between the present approach and that of [23] and [24]. In the latter references there is a strict adherence to hierarchical processing. That is not the situation here, where there is a great deal of lateral connectivity in the system. One may consider the net D as at the top of a hierarchy, but there is only at most one hierarchy in the system.

The basic circuit arrived at is shown in figure 3. This extends the circuit of figure 2 by the addition of an input processing module I and feedback lines from the decision unit D to I. The module I can be the primary visual cortex, areas 17 and 18, in the visual modality, whilst the presence of feedback lines is well established. One of the functions of the feedback lines, as considered extensively in [23] and [24] is to control the primary inputs. Thus there may be inputs which lead to feedback from D which cause a reduction in the further primary inputs. The ensuing activity of the system is then controlled by internal activity circulating round the internal loops which are both explicitly drawn in figure 3 ($W \rightarrow D \rightarrow I \rightarrow W$, $W \rightarrow E \rightarrow D \rightarrow I \rightarrow W$) and implicitly ($W \rightarrow E \rightarrow W$, $W \rightarrow D \rightarrow W$, etc).

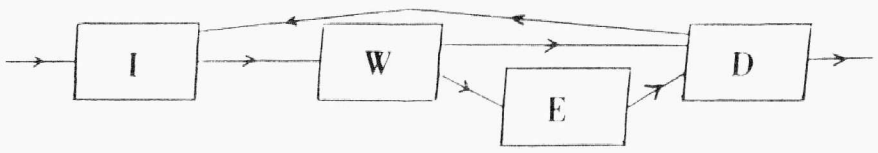
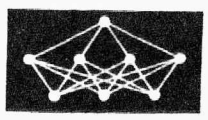


Fig. 3. Extension of the machine in the previous figure to include the preprocessing unit I and the feedback from the decision unit D to I to achieve control of the input, as discussed in the text.

It has been claimed above that learning by means of temporal topographic maps will lead to causal connections of both an intra- and inter-module form which encode the causal structure of the inputs being experienced. The resulting activity when the external inputs are inhibited, as described above, should thus develop along the same lines as it did when there was no such inhibition. In particular there will be the development of logical chains of thought, arising by the interactions between modules in W or E. There will also be modifications of these causally guided developments of activity due to the altered input set. Thus the nature of the experience of the system in such an altered state will be, in general, of a considerably different form than that in the originally aware state. It would indeed be appropriate to call such a state an altered state of consciousness.

It is clearly possible to include in the model further states of consciousness, such as REM or slow-wave sleep. That has already been developed in the references [23] and [24]. We do not wish to do that here at this juncture, since it does not seem crucial to our modelling purpose.

We conclude finally that the system is indulging in autistic thought in the above activity. It may well be described as day-dreaming.



6. Conclusions.

We have presented a general blueprint for a machine which can, with some reason, be said to think. Yet there are clearly some very difficult problems we have to solve before we can put the system into action so as to test its powers.

Firstly the temporal topographic map learning algorithm must be tested by simulation. That can be done straightforwardly for small nets, with small input sets, and is presently under investigation. This is expected to lead to a useful set of nets trained on simple (small) tasks. But will these nets scale in any effective manner? We raised this question earlier from a technical point of view. However here it is necessary to consider this point from a general viewpoint. Is it the case that the above approach has not yet got the crucial clues as to the way that neural nets work at the level of efficiency of our own brains. Are there still subtleties of neurons, of synapses, or of other aspects of living neural nets not yet included in our discussions which are essential for the success of the enterprise of building thinking neural nets?

We think not. The subtleties of the above sort left out so far in our discussion undoubtedly do help the system. That has already been noted in the analysis of temporal sequence storage in [10]. But the experience gained there in the need, for example, for a description of the opening and closing of synaptic channel gates, indicates that such details improve the system performance but do not change the underlying modes of operation. Thus it seems reasonable to claim that the complete description of the manner of operation of neurons, synapses, etc., down to the molecular level will be a technical aid to the system in allowing it to be scaled up without drastic loss of efficiency, but need not alter in any way the general algorithms under which the system is operating.

The second area of discussions is that of a very different sort. It is as to the nature of the experience of the machine when it is in action. Will it, in particular, have a sense of conscious awareness that we, as humans, all experience and which has led to the mind/ /body problem? That is not a question which we will discuss in any detail here, but note that the system is to be expected to develop a model of its own actions. In this way it will develop a sense of consciousness of self. The manner in which this could be simulated depends on the success of the earlier simulations of the nets *W* nad *E*. We have a long way to go before we can begin to appreciate the problems raised by such features.

References

[1] J. Searle: Minds, brains and programs, Behavioral and Brain Sciences, Vol. 3, 1980, 412—457.

[2] P. Churchland: Neurophilosophy, Bradford Books, 1986.

[3] J. E. Fenstadt, J. V. Benthem and T. Loenning: Situations, Logic and Language, Reidel Publ. 1987.

[4] J. A. Fodor and Z. W. Pylyshyn: Connectionism and cognitive architecture: A critical analysis, Cognition, Vol. 28, 1988, 3—71.

[5] A. Clark: In defence of Explicit Rules, in: Philosophy and Connectionist Theory, ed. Ramsey, Rumelhart and Stich, 1989.

[6] I. Croall: Report on ANNIE, Mason Conf. on Neural Networks, BAAS, Swansea, 1990.

[7] J. G. Taylor: Living Neural Nets p. 31—52, in: "New Developments in Neural Computing", ed. J. G. Taylor and C. L. T. Mannion, Adam Hilger, 1989.

[8] J. G. Taylor: Spontaneous behaviour in neural networks, J. Theor. Biol. Vol. 36, 1972, 513—528.

[9] P. C. Bressloff and J. G. Taylor: Random Iterative Networks, Phys. Rev. Vol. 41, 1126—1137, 1990; D. Gorse and J. G. Taylor, A general model of stochastic neural processing, Biol. Cyb. Vol. 63, 299—306, 1990.

[10] M. Reiss and J. G. Taylor: Storing Temporal Sequences, KCL preprint, 1990.

[11] K. J. Stratford, A. J. Mason, A. U. Larkman, G. Major and J. J. Jack, in: The Computing Neurone, ed. R. Miall and G. Mitchison, Addison-Wesley, 1989.

[12] K. A. Martin: From single cells to single circuits in the cerebral cortex, Quart. J. Physiol., Vol. 73, 1988, 637—702.

[13] A. K. Engel, P. Konig, C. M. Gray and W. Simger: Stimulus dependent oscillations in cat visual cortex, Eur. J. Neurosci., Vol. 2, 1990, 588—606.

[14] E. S. Spelke: Where perceiving ends and thinking begins, p. 197—234, Perceptual Development in Infancy, vol. 20, ed. A. Yonas, Erlbaum, 1988.

[15] T. Kohonen: Associative Memory, Springer, 1984.

[16] W. Penfield et al: Speech and Brain Mechanisms, Princeton Univ. Press, 1959.

[17] J. G. Taylor and M. Reiss: Does the hippocampus store temporal sequences?, KCL preprint, 1990.

[18] G. Z. Sun, H. H.-Chen, C. L. Giles, Y. C. Lee and D. Chen: Connectionist Pushdown Automata that learn Context-Free Grammars, in: Proc. IJCNN-90, Wash. DC, ed. M. Caudill, Erlbaum Assoc. Hillsdale, N. J., 1990, pp. I—557—580.

[19] G. Major, A. Larkman and J. Jack: Constraining non-uniqueness in passive electrical models of cortical pyramidal neurones, Proc. Physiol. Soc. Vol. 23P, 1990.

[20] J. G. Taylor: in preparation.

[21] A. Waibel: Modular Construction of Time-Delay Neural Networks for Speech Recognition, Neural Comp., Vol. 1., 1989, 39—46.

[22] J. G. Taylor: in preparation.

[23] W. T. Powers: Behaviour: The Control of Perception, Wildwood House, 1973.

[24] W. Stallings: A. Cybernetic Model of States of Consciousness, Kybernetes, Vol. 4, 1975, 225—231.

ASSOCIATIVE INTERNEURONAL BIOLOGICAL MECHANISM

J. Faber)*

Abstract:

The neurologist finds analogies between the Farley and Clark automatic self-organizing model and the brain highly intriguing. The signal generator suggests comparison with the thalamus which also has a rhythm-making function and, likewise, sends many variables — impulses — into the cortex. The complex with its elements randomly connected at the start of the experiment is reminiscent of the cortex which, in the newborn, is in a naive, poorly organized state. The discrimination unit designed to determine the state values of the cortex is like the limbic system which monitors theh body's metabolic equilibrium by means of internal environment receptors in the hypothalamus, and which adjusts the "emotive equilibrium of mental functions" by means of endocrine and nervous mechanisms. Stimuli from the discrimination unit travel on to the signal generator and to the formator. The formator can be likened to the modulatory humorergic centres it similarly regulates the thresholds of elements and connections in the cortex and other parts of the brain. In the model there is one formator, in the brain there are more. For each state there is a centre of formator action: the reticular formation for the state of vigilance, nuclei raphe for synchronous sleep, locus caeruleus for paradoxical sleep. Each nucleus operates in its own way, generally perhaps by setting the threshold and, consequently, by changing the programmes of the target neuronal circuits and networks. Under pathological circumstances, even a cortical lesion, e.g. an epileptic focus, can become a formator. This focus then competes with physiological formators for control of the cortex. This power struggle then results in an epileptic attack or acute psychosis. For the most part, physiological formators act as inhibitors. During epileptogenesis, prior to manifest paroxysms, there is gradual loss of sleep, especially paradoxical sleep.

1. Introduction

Without making any megalomaniac claims to explain the activity of the brain, some of the similarities between artificial cybernetic systems and brain functions are so striking that we cannot help drawing some analogies in terms of function as well as structure.

*) Prof. J. Faber,
Department of Neurology
Charles University,
120 00 Prague 2, Kateřinská 30
Czechoslovakia

2. The Farley and Clark model

The auto-organization system of Farley and Clark (1954) is one of the first and, to this day, in many respects unsurpassed model of brain function *(Fig. 1)*. Designing the model, the authors used four principal components: signal generator, complex, discrimination unit, and formator.

2.A. The signal generator produces sequences of signal, i.e. two different, though regular, sequences to send them on to the complex. The generator itself receives control signals from the discrimination unit.

2.B The complex consists of 128 initially randomly connected elements, whose mutual contact undergoes no further change during the experiment. These elements are divided into 4 fields, ie. 2 input and 2 output fields. The sequences of signals from the generator reach the two input sectors (01 and 02) and from there pass on quite randomly to the two output sectors (O+ and O−). The point is that one type of the sequence of signals should arouse response solely in the O+ sector and the other solely in the O− sector. This "spontaneous" discrimination is effected so that the initially random connections between the elements of the complex become organized, i.e., some connections acquire a low threshold of excitability and become "important" while others become extinct. For this decrease in entropy among the 128 elements to come about sufficiently soon, the system needs two more members: the discrimination unit and the formator. There are, however, self-organizing autonomous complexes of elements without any such "accessories", but then the elements must be endowed with a fairly large independent memory: even so, however, the "learning" period is longer than in the Farley and Clark model (Beneš 1966).

2.C. The discrimination or analytical unit analyzes the state of the complex, i.e., by monitoring the state values which show how the complex has "learned" to discriminate between the two types of sequences of signals. The information about the state of the complex organization is then passed on to the signal generator, thus influencing the alternation and number of one or the other sequence of signals. In the other route, information about the state of the complex is channelled into the formator.

2.D. The formator or modifier is the last member of the system. As already said, it receives information from the discrimination unit and, on their evaluation, sends out control signals or action quantities into the

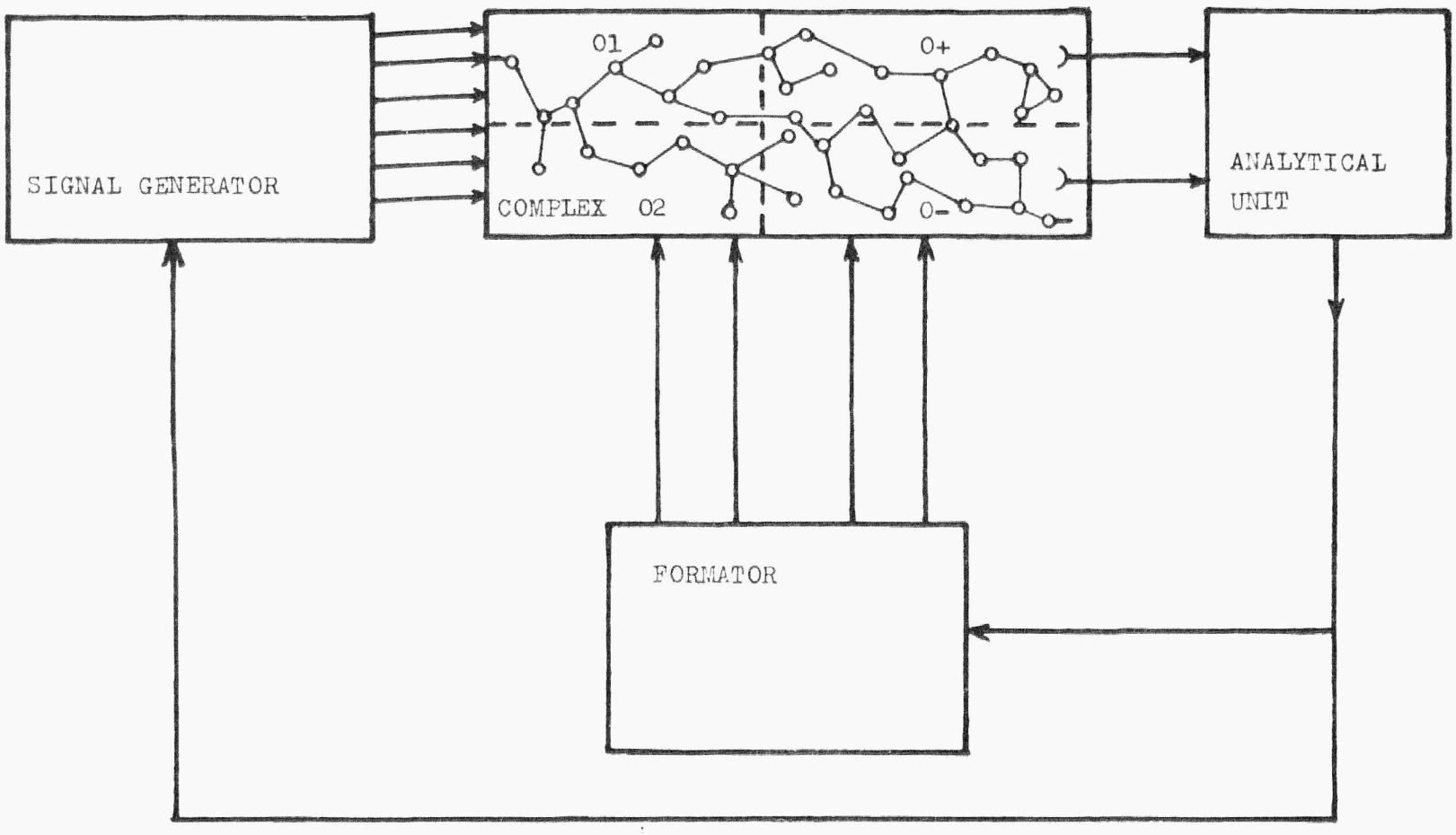


Fig. 1. Scheme of the Farley and Clark model.

complex. The formator effect on the complex is, essentially, of dual type: it controls the thresholds of excitability of the elements of the complex as well as the capacity or conductivity of the inter-element pathways, and second, it transmits Gaussian noise into the complex. This noise randomizes the deterministic process going on in the complex. Thus the process becomes less inhibited but freer in looking for the way to a solution, which, oddly enough, accelerates the business of "learning" and seeking an equilibrium.

Noise has a major significance in the process of automatic organization as H. von Foerster showed (1960, according to Beneš — 1966). Nicolis et al. (1975) refer to the significance of different types of noise. These authors employed a mathematical model of a self-organizing "open non-linear system remote from thermodynamic equilibrium" to realize it on a computer. As they exposed the process of organization to intensive noise which was stationary in terms of amplitude, the organization of the system was growing, entropy decreasing and redundancy increasing. However, when they used "jitter" noise which was non-stationary in terms of amplitude, the organization was being retarded or else there was an increase in entropy.

We ourselves studied the significance of noise as a regular variable carrier to find out that 4 % of the harmonic variable in the Gaussian distribution of the random variable was sufficient for this regular signal to be detected by the periodogram. In contrast, in the Weibulian distribution of noise, 8 % of the harmonic variable was needed for the detection (Faber and Vladyka 1984).

Beneš developed the theory of the formator control of the complex to apply it, for instance, for the control of large systems. He also devised a model of the effect of the human factor in large systems (Beneš 1981, 1990). Kotek et al. (1980) used a somewhat remote mathematical apparatus for a similar function, i.e., adaptation and learning.

3. The model — brain analogy

The properties of the Farley and Clark model and, in particular, those of its individual parts are very interesting in that they suggest comparison with different parts of the brain *(Fig. 2)*.

3.A. The generator of signals can be likened to the thalamus, which is the largest nucleus or grouping of neurons in the highest portion of the brain stem. This structure is noted, in particular, for two important properties: 1. it is the last transmitter of impulses from the peripheral sensory organs such as the eye, the ear, and cutaneous, muscular, articular and tendinous receptors, impulses which, on processing, are passed on to the cortex; 2. it is a generator of rhythms, to which impulses from the periphery of the body are also partially subordinated. Hence it also has a clock pulse function. Moreover, in the thalamus all the nuclei representing different sensory organs are abundantly interconnected, and exercise an indirect effect on the power of movement. This is probably the site of a primitive integration of impulses and, thereby, of the subject's simple awareness of its body and the environment. This is definitely the case of beast of prey, though a similar phenomenon cannot be ruled out in man either. For instance, even with the visual cortex removed, the dog can move about without knocking against surrounding objects (Babák 1908). The thalamus is sometimes referred to as the gateway of consciousness *(Fig. 3)*.

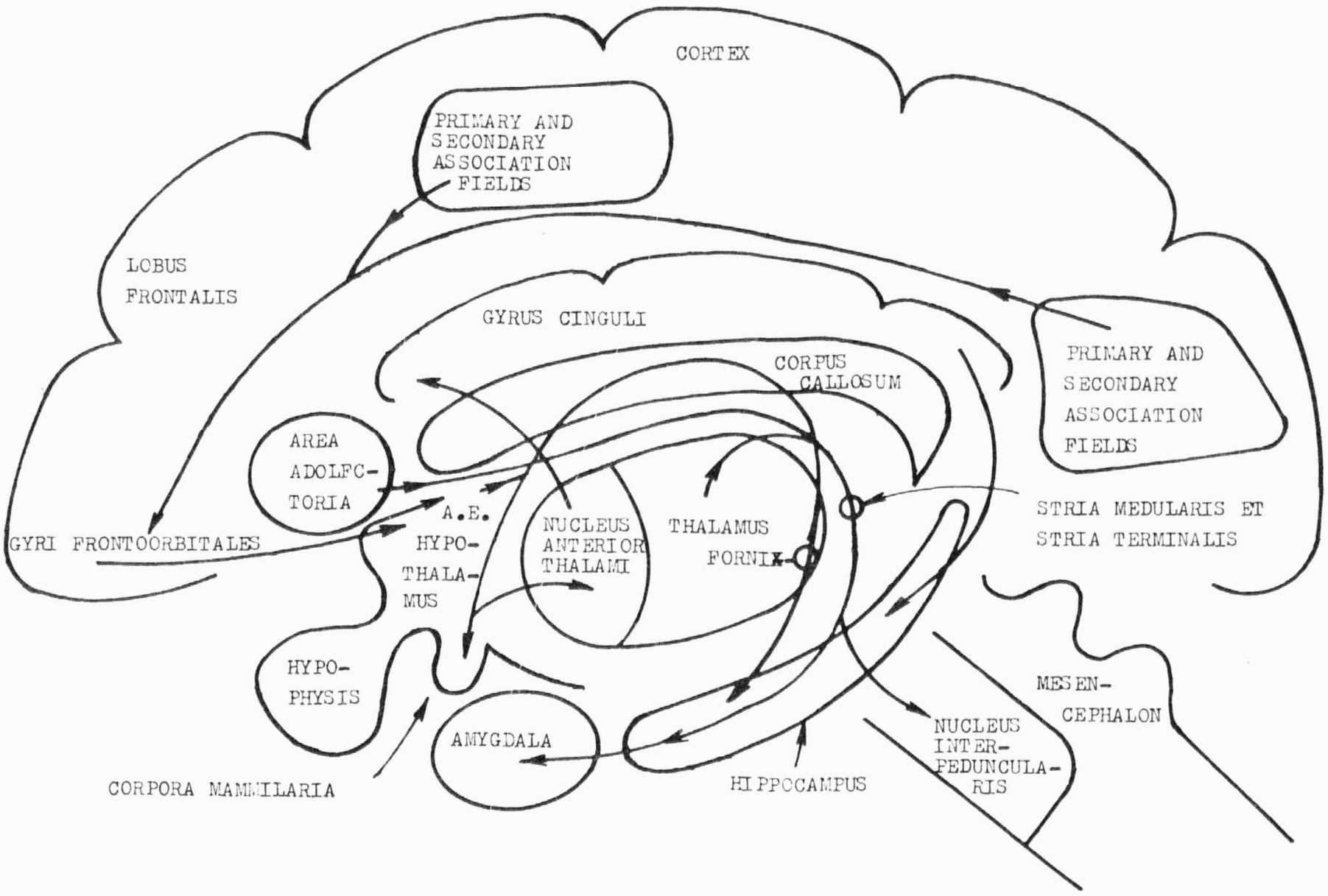


Fig. 2. Semischematic view of the sagital section of the brain. A. E. is area entorhinalis (area 28), part of the palleocortex.

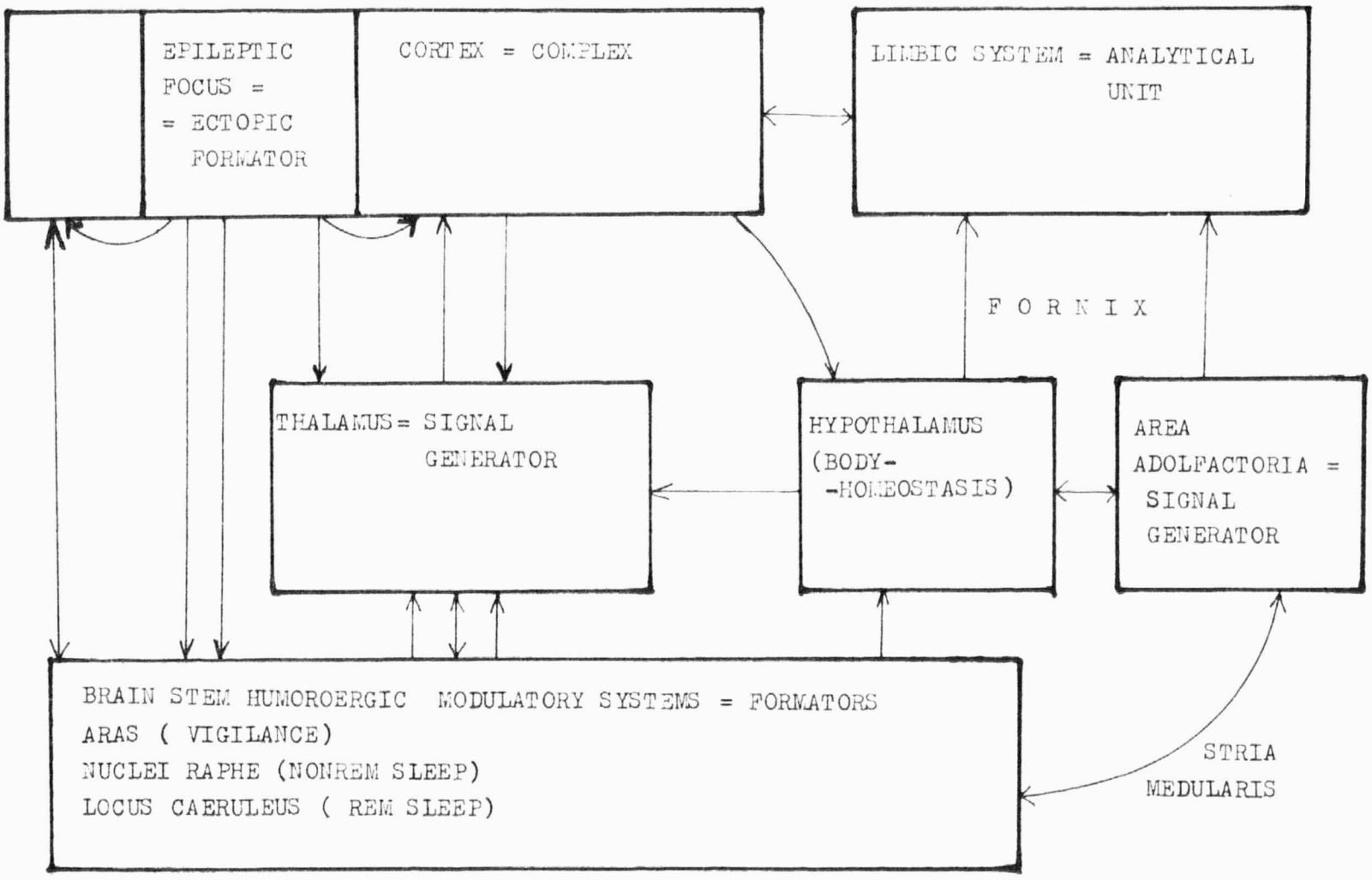


Fig. 3. Analogy between brain structures and the Farley and Clark model.

3.B. The complex with its initially randomly interconnected elements reminds us of the neonatal cortex with its naive "unenlightened" neurons. It is only upbringing and education that will organize the brain. With regard to the model, however, there is a measure of difference, e.g., in that the neurons of the cortex are not quite so randomly interconnected as their processes, dendrites and neurites grow in a partly organized way in accordance with genetic rules, i.e. by giving rise to the left and right hemispheres, in each hemisphere to four lobes (frontal, parietal, temporal and occipital), according to Brodmann to 50 areas and, according to Hubel and Wiesel (1962), to some million columns. Each column contains about ten thousand neurons, thus representing a group of neurons organized in a column which runs vertically right through the cortex. The total number of neurons in the human cortex is about 2.10^{10}. Also genetically developed is the six-layer structure of the cortex and some of the connections between the columns. Terminologically speaking, this is the neocortex which accounts for 95 % of the surface of the cortex. Between the thalamus and the cortex, impulses keep circling reverberating, hence we refer to the thalamo-cortical reverberation circuit (TCRC).

If anything bars the physiological flow of impulses from the eye to the cortex, the genetically started columnar organization will not reach completion. A similar situation occurs in individuals born blind who even fail to develop the appropriate electroencephalographic (EEG) activity of the brain marked by alpha activity at a frequency of some 10 Hz (Cohen 1969, Hubel and Wiesel 1962). EEG activity takes some 5 to 6 years to achieve maturity. In the school age it is already like that in adulthood. A number of animal experiments have shown the relationship between the genetic plan and postnatal programming. As Valverde (1971) showed, mice kept in darkness have porly developed dendrites, spines and synapses of neurons of the visual cortex while animals living in daylight have visual cortex neurons rich in those organelles. In other words, an adequate influx of impulses to the cortex accounts for the development of plentiful interneuronal connections.

We can also put it in this way: in a living brain "frequently used software will change into hardware" (Faber and Vladyka 1987, Faber and Weinberger 1988). Hence also in the treatment of psychoses it is advisable to use "kind words" all the time as the principal programming instrument, and not only chemotherapy. Summing up years of experience, Sjöström (1985) noted that his schizophrenics receiving psychotherapy felt subjectively better and objectively exhibited greater sociability on substantially lower drug doses than patients left to pharmacotherapy alone.

Human pathophysiology recognizes atrophy of the cortex, e.g. Alzheimer's or Pick's diseases, conditions invariably associated with dementia and personality disintegration. Consequently, the cortex is the seat of intellectual abilities and probably also the seat of memory engrams. The brain has many inborn reflexes (sucking, grasping and other reflexes) with centres situated subcortically, but relatively speaking, with regard to adulthood, the cortex is a "tabula rasa". The huge number of neurons appears to be necessary only in the actual phase of learning. Once the perception of and motor response to stimuli have become automatic, the need for the number of neurons is substantially less. Similarly, vigilance in itself utilizes few neurons. Thanks to the great redundance of neurons, "any kind of programme can be implemented" in the cortex, as witnessed by deprivation experiments in animals or the unfortunate stories of humans such as, e.g., "wolf children" or the fortunes of Kašpar Hausner (Veselovský 1985, Říčan 1975). This means that a sound brain can learn anything from the ability to live in the jungle up to the potential career of scientists. One cannot help wondering about the number of partially deprived "modern, TV-educated children".

3.C. Another member of the automatic self-organizing system, the discrimination unit, can be likened to the limbic system (LS). Phylogenetically, this is one of the oldest parts of the brain, i.e., the paleo- et archicortex; in pre-mammalian vertebrates closely connected with the smell brain or rhinencephalon, and, in those species, hierarchically standing at the top of the central nervous system (CNS).

The limbic system receives information from the whole body, mainly from visceral organs, and also controls those organs by way of the hypothalomo-pituitary system and vegetative nerves. The hypothalamus, which represents 0.5 % of the weight of the brain, registers and regulates body temperature, osmotic blood pressure, blood sugar levels, the acid-base balance of the blood, and it takes a share in sleep regulation (nucleus suprachiasmaticus). The LS also receives information from the neocortex, from the so-called primary and secondary association fields, i.e., information mediated from the periphery, from the sensory organs (eye, ear, muscle receptors, etc.).

The LS includes the septal nuclei, in humans — the area adolfactoria representing a clock pulse generator specially developed for some parts of the LS, mainly for the hippocampus. Phylogenetically, this structure, the hippocampus, is the old cortex — the archicortex. It has a mere three layers of cells and, together with another part of the cortex (gyrus cinguli) occupies less than 4 % of the cortical surface. The remaining one per cent is the oldest part of the cortex, the paleocortex, i.e., cortex prepiriformis and part of the amygdala. The hippocampus with the adjacent structure, the amygdala, are important centres for olfactory perception and for the rise of emotions. According to Gastaut (1952), this is the site of origin of the stress reaction advancing to the hypothalamus, the pituitary, and the adrenals which produces the hormones responsible for the body's "sympatheticotonic" reaction involving heightened blood pressure, quickened heart rate, dilatation of eyes and pupils, and subjective,

mostly negative, feelings, such as anger and susceptibility to aggressiveness.

The hippocampus is the site of a kind of time register designed to classify memory traces according to "when and where they took origin" (O'Keefe and Nadel 1978). This is an invariably affect-dependent process. Like all mental activity, committing something to one's memory, remembering, is always associated with some kind of emotion. The stronger (and more positive?) the emotion is, the better we can remember. We learn because we are motivated, because we take a certain stand to learning. Consequently, the LS acts to combine rational and emotional behaviour, thus helping to create the personality structure.

Motives can be classified differently. For instance, we refer to the biological and social origins of motives. Primary or biological motives include, for instance, instincts as complex inborn unconditioned reflexes (e.g., birds building nests in the spring, exhibiting parental instincts, procreating the young, flocking and flying away in the autumn). In humans, such motives are called impulses. (The word impulse has a dual meaning; the above mentioned type of motives, and the electrochemical process of transmitting a stimulus along the membranes of dendrites and neurites from one neuron to another). Impulses represent endogenous character-based tendencies. There are, for example, individuals of moderate or energetic behaviour; in Hippocratic terms — people of sanguine, choleric, phlegmatic or melancholic temperament. We can refer to a mental cast with a propensity to rational or artistic thinking, hypersexuality, toxicomania, etc. (Madsen 1972). A number of psychologists have devised ingenious qualitative and quantitative criteria for the description of human nature. Thus, Eysenck classifies differences in personality by the degree of neuroticism (the polar extremes being suppressors and sensitors), and on the plane vertical to the previous characteristics he makes a distinction between the degrees of "sociability" (the polar extremes being extroverts and introverts). *(Fig. 4).*

Besides impulses there are incentives, i.e., parts of motivation induced by a certain type of education, especially in early childhood, mostly within the family. In ideal cases, impulses and incentives work hand in hand. For example, the child has a talent for music, and one of the parents is a musician. Thus the child is raised in a harmonious relation of impulses and incentives, i.e., learning music, enjoying it and scoring successes. Naturally, there are many degrees of transition between the ideal and the pessimal such as when a pubescent is impulsively (instinctively) hypersexual while education in the family is marked by strong religious or puritan leanings. The youngster's further progress then depends on his or her other characteristics of nature. In any case, however, a conflict arises due to the discrepancy between these two lines of motivation. Unless the motivated behaviour is gratified,

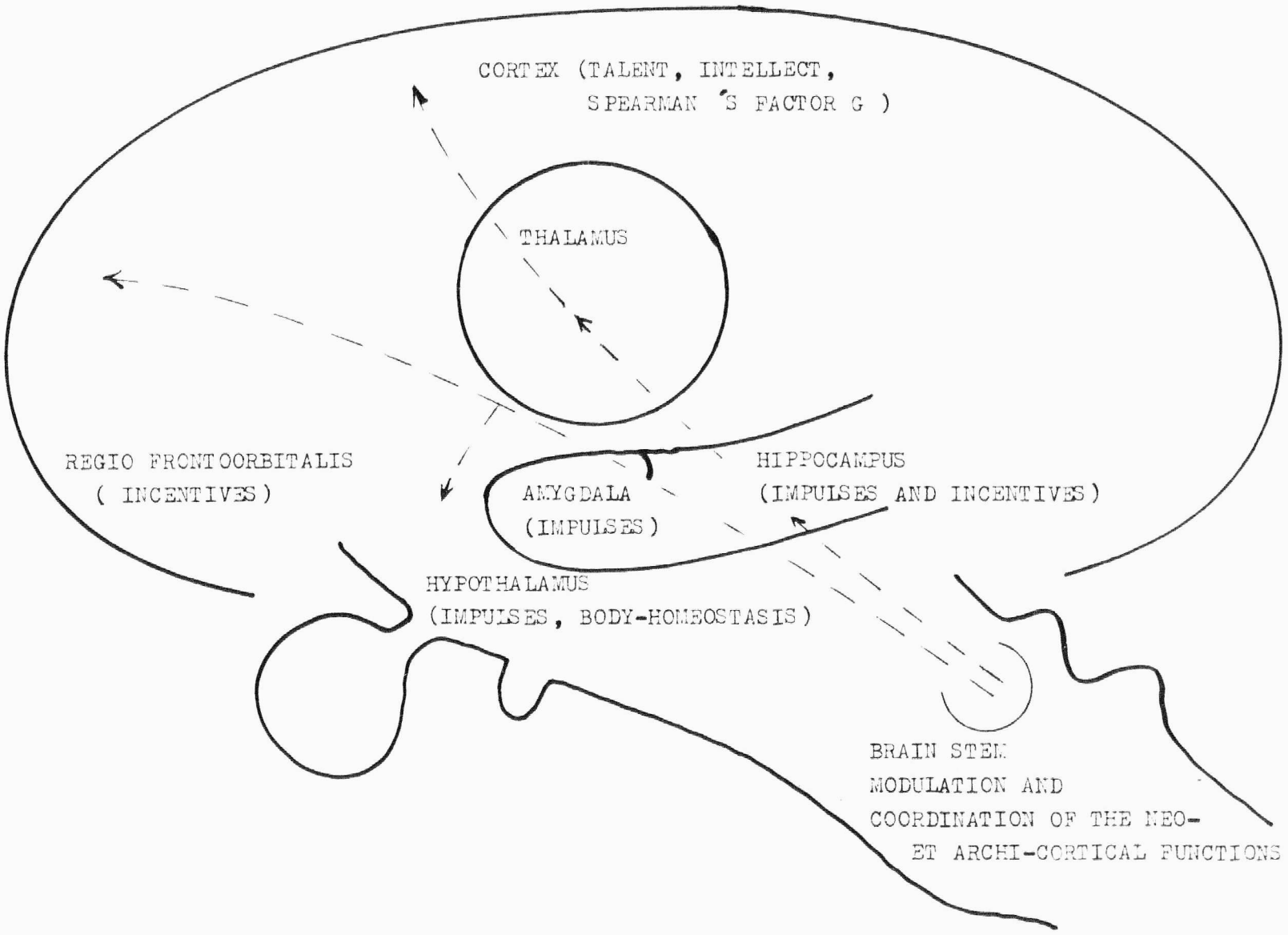


Fig. 4. An attempt at localizing the motives in the brain.

there is bound to be frustration or even total deprivation of this instinctive need. Depending on their nature, the individual may respond to this in many ways, e.g., by a show of depression or regression (infantilism) or aggression, mostly, however, in the presence of subjectively unpleasant emotions and, objectively, often with signs of stress.

This is the classic way to neurosis arising from a conflict of motives. The whole process may take an entirely unconscious course. A good psychoanalyst can identify the presence of such unapparent conflicts and offer the patient relief by helping him to expose such latent psychotraumata. Together, they can then try and evoke the traumatizing situation with the patient fully aware of all the circumstances and, eventually, alter the patient's insight and help him get rid of the cause of chronic tension, anxiety or somatic symptoms (palpitation, variable hypertension, gastrointestinal ulceration, etc.).

The human brain is endowed with an immense power of memory; there is a daily build-up of experiences, and it is quite certain that not all of them can be remembered equally well so as to be recalled at any time. Moreover, unpleasant experiences are often well remembered but recalled with subconscious reluctance as there is an emotive block. Their persistent subconscious presence may become even more harmful as they take up too much operation time in the CNS. Like the discrimination unit in the model, the LS in the brain keeps monitoring the state of the viscera (cardiovascular and gastrointestinal organs), the body (receptors of the skin, muscles, joints) and the neocortical — rational situation, i.e., the actual picture of the animal's situation in natural surroundings and man's in society. It is the LS that keeps all this mutually integrated, thus creating the momentary psychosomatic situation described as "mood" or affectivity (Plzák 1975). This may be either positive (joy, cheerfulness) or negative (sadness, anger). There is always a vegetative reaction (e.g. heightened heart and respiratory rates) and secretion of hormones: from the pituitary gland — prolactin, STH (growth hormone), ACTH (adrenocorticotropic hormone) and, as a secondary source, adrenaline and corticoids from the adrenals. Studying epileptics with implanted electrodes, Gallagher (1987) found that hippocampal epileptic discharges shorter than 10 seconds inhibited the hypothalamo-pituitary system as well as ACTH secretion, while those of longer than 10 second duration increased the secretion of ACTH and prolactin. This clinical fact serves as a model of the LS effect on stress and mood in terms of both excitation and inhibition.

The LS system appears to take a major share in a phenomenon which can be described as "awareness of reality". The LS evaluates interoceptive (visceral), somatic (skin, muscles) ad exteroceptive (eye, ear) events, whose equilibrium gives rise to the emotive charge and "somatization" of that which is experienced. Under pathological circumstances, the limbic system is susceptible to irritation leading to epileptic psychomotor paroxysms or to brief psychotiform states. A paroxysm is often preceded by the sudden appearance of an aura of a sensory or emotive nature, e.g., pseudohallucinations of odours and anxiety. This is not loss of consciousness, but rather loss of a function of memory with the patient losing real contact with the surroundings and lapsing into automatic non-adequate behaviour. This kind of situation can repeatedly be observed in epileptics and psychotics with electrodes implanted into the amygdalo-hippocampal complex (AHC) during spontaneous and provoked epileptic discharges. High sharp and slow waves are discernible in the AHC itself while the neocortical system may show near normal activity.

The TCRC has the following principal functions: keeping lucid vigilance, i.e., reactivity, perception and adequate motor response to stimuli, in other words, gnoseological and psychomotor functions. In this kind of system, epileptic discharges cause loss of consciousness, i.e. uncounsciousness. What subsequently appears on the EEG are regular S+W complexes, i.e. spikes and waves taking turns. At the same time, the LS may be little affected by epileptic activity. The LS is prominntly active during paradoxical sleep, hence our tendency to mistake dream hallucinations for reality.

3.D. The last part of the system, the formator, reminds us of the activity of the brainstem modulation humorergic centres. As the brain is a far more complex structure than a cybernetic model, it comprises more such formators. As it helps to maintain vigilance and sleep, it also serves, albeit indirectly, memory and learning. Information storage in the memory medium is relatively simpler in a non-living machine than in neurons. The computer will process data only if they are in a logical-arithmetic unit while the brain appears to process such data more often, and that kind of processing unit (neuronal network) cannot well be distinguished from the memory medium. To put it in a simplified form, memory is in every neuron and every neuron is a small processor.

The activation reticular ascending system (ARAS) is localized in the pons and in the mesencephalon and its purpose is to maintain vigilance. When Fréderic Bremer (1935) surgically interfered with those structures in the cat, the result was unconsciousness. As Moruzzi and Magoun (1949) discovered, it was not the whole mesencephalon that was affected but merely its central portion, the so called tegmentum, whose neurons stimulate with their axons the thalamus and, indirectly, the cortex, thus keeping the animal awake. Hence, the ARAS is the formator for vigilance.

Formators periodically take turns in their activities. After a certain period of wakefulness, say after 16 hours, control of the brain is taken over by centres for synchronous or NONREM sleep. They are nuclei of the medulla oblongata, nuclei raphe, nucleus tractus solitarii, and nucleus suprachiasmaticus hypothalami secreting serotonin. A share in sleep regulation is also

taken by other, as yet unidentified, peptidergic centres secreting proteins such as VIAP (vasoactive intestinal peptide) and DSIP (delta sleep inducing peptide) which is found in the cerebrospinal fluid of the ventricles of brain and which, if supplied into the ventricles, even in its synthetic form, or into the blood of animals, will induce sleep (Monier and Schöneberger 1976).

At this point it is useful to realize that to model human nervous and psychic activities it is necessary to monitor not only the neuronal networks but also other, e.g., humoral modes of information transfer with their effect on the chemical nature of impulses.

The prelude of synchronous sleep lasting about two hours is followed by a period of paradoxical or REM (rapid eye movements being typical of this period) sleep.The nuclei (i.e. anatomically defined groups of neurons) reponsible for REM sleep are localized again in the brain stem (being its formators): locus caeruleus in the pons Varoli and ncl. gigantocellularis. Similarly to the previous structures designed to regulate wakefulness or REM sleep, these nuclei send out their fibres (neurites) over long distances to other parts of the brain stem, the spinal cord, hypothalamus, basal ganglia (where they programme and regulate skeletal muscle movement), thalamus and cortex. They take their name from the substance (modulator) which they secrete at their synapses. Thus, the ARAS is associated with acetylcholine, NONREM with serotonin, dopamine and peptides, REM with noradrenaline and again acetylcholine and peptides. Hence also the names of acetylcholinergic, serotoninergic and other systems *(Fig. 5)*.

During REM sleep, the blood receives endocrinological secretions of the hormones prolactin, testosterone, endorphins and, imprecisely synchronously, also ACTH. During NONREM sleep, the STH (growth hormone) is secreted. Generally speaking, melatonin and, in the last third of the night, ACTH and cortisol are secreted during sleep. Paradoxical sleep is marked by increased activities of neurons of the cortex and the whole motor system including the cerebellum, brain stem and spinal cord. However, the motor pathways are blocked at the level of motoneurons of the spinal cord, which is why no living being can move about in REM sleep. In 1965, however, Jouvet and Delorme made a surgical operation on the distal part of the locus caeruleus so that their cat after the operation performed with its movements what it was dreaming about. What apparently applies to all mammals is that what we experience in dreams goes on not only on the sensory, i.e. imaginative level, but also on the motor level. That means: in a dream involving swimming or running our brain does, indeed, produce specific movements for swimming or running, and it is only the block of motor pathways as a last resort in the spinal cord that prevents us from actually realizing the movements *(Fig. 6)*.

"Jouvet's cat" permits us to explain some of the syndromes of the human pathology of sleep such as, for example, somnambulism (sleep walking) or pavor nocturnus (nightmare) and perhaps also some types of sleep psychomotor epilepsy. The busy neuronal activity goes hand in hand with intensive mental activity in the form of dreaming. REM dreams are very vivid, made up of all the sensory qualities and noted for their emotive charge. According to American statistics, two thirds of what we experience in our dreams

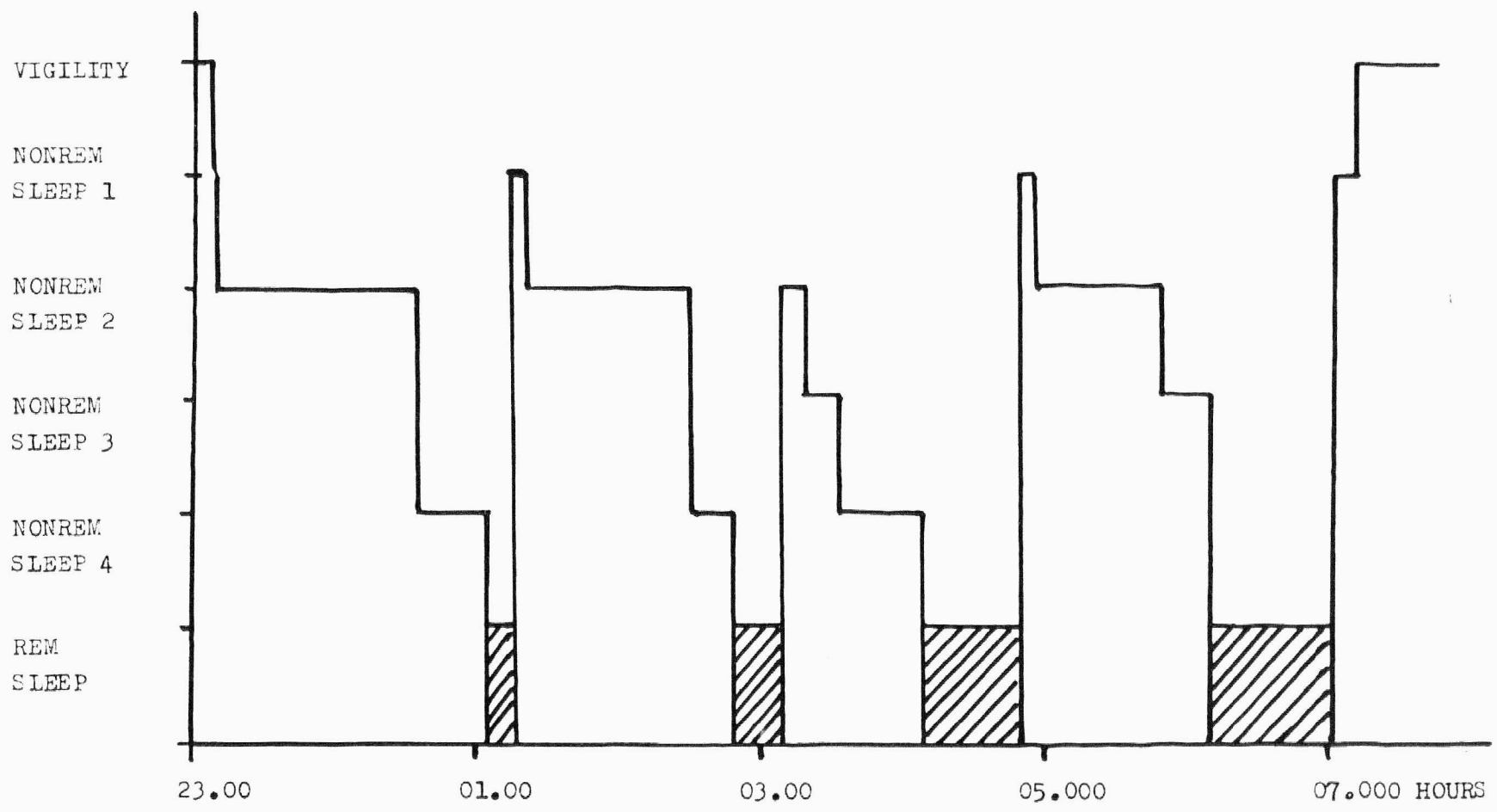


Fig. 5. An example of typical hypnogram, i.e. graphic representation of normal changes of NONREM and REM sleep phases.

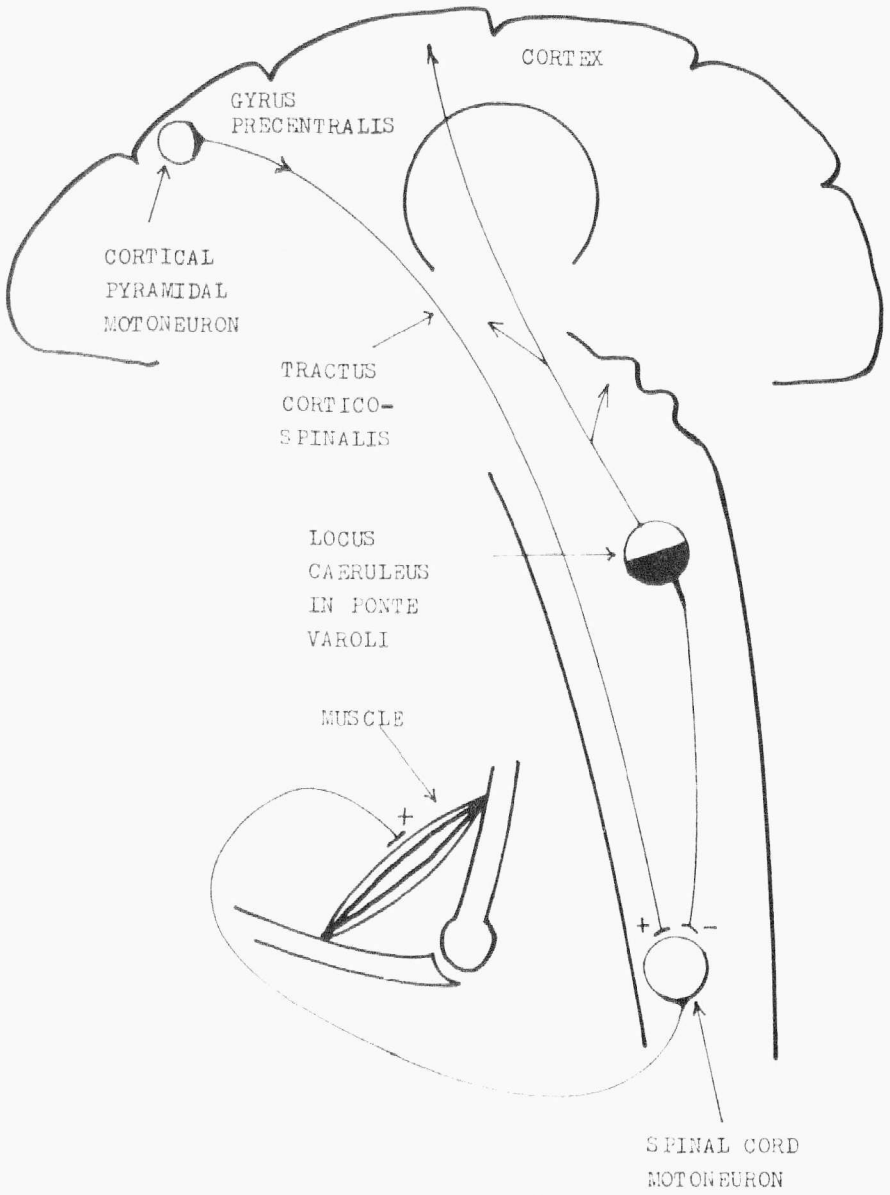


Fig. 6. A representation of Jouvet's experiment with desinhibition of the spinal cord motor mechanism during REM sleep. Distal part of the locus caeruleus (solid) is destroyed.

are negative emotions. The motor pathways block is manifested as atonia, i.e. loss of normal muscle tone, which appears to account for reflex dreams about falling or flying.

Different states of consciousness are matched by different EEG rhythms: wakefulness is noted for alpha activity (8—13 Hz), NONREAM sleep for theta (4—7 Hz) and delta activities (1—3 Hz) of higher amplitude and for sleep spindles (14 Hz), and paradoxical or REM sleep for theta, delta and beta (14—40 Hz) of low amplitude. During sleep, neurons undergo major metabolic changes: NONREM is marked by a rise in RNA, and REM by an increase in the level of proteins which seem to be the substrate of memory (Hydén 1980). During NONREM sleep, some neurons slow down, others speed up the rate of discharges while during REM sleep, all systems of the brain, sensory and motor areas and their neurons accelerate their activities (Hobson and McCarley 1971, (Faber 1978).

What remains a moot point is the way the brain stem nuclei as formators control different states of consciousness. No doubt, this involves adjustment of the input and output thresholds of neurons, an operation performed, in principle, by synapses of dual type: axosomatic synapses secreting mediators of fast and short-term effect on membrane polarization (e.g., GABA, glycine), and axodendritic synapses secreting modulators of slow and long-lasting effect on the membranes (e.g., noradrenaline, dopamine). *(Fig. 7).* Between the thalamus and the cortex there are milions of fibres with thousands of millions of impulses per second circulating in them. For a system like that to be able to operate effectively, it has to meet a number of conditions:

1. The number of impulses per second has to be optimal; a very low number is present during unconsciousness, e.g., due to asphyxia, and very high during an epileptic attack.
2. The same goes for opptimal synchronization, i.e., the simultaneous co-operation of many columns of the cortex and rhythmic generators of the thalamus.
3. The sequence of impulses must be meaningful, i.e., it must carry some encoded information. This is a phenomenon which Mountcastle (1966) called "neural replica". The point is that a certain "firing pattern" of impulses in the neuron always represents a piece of specific information, e.g., the pitch of a tone

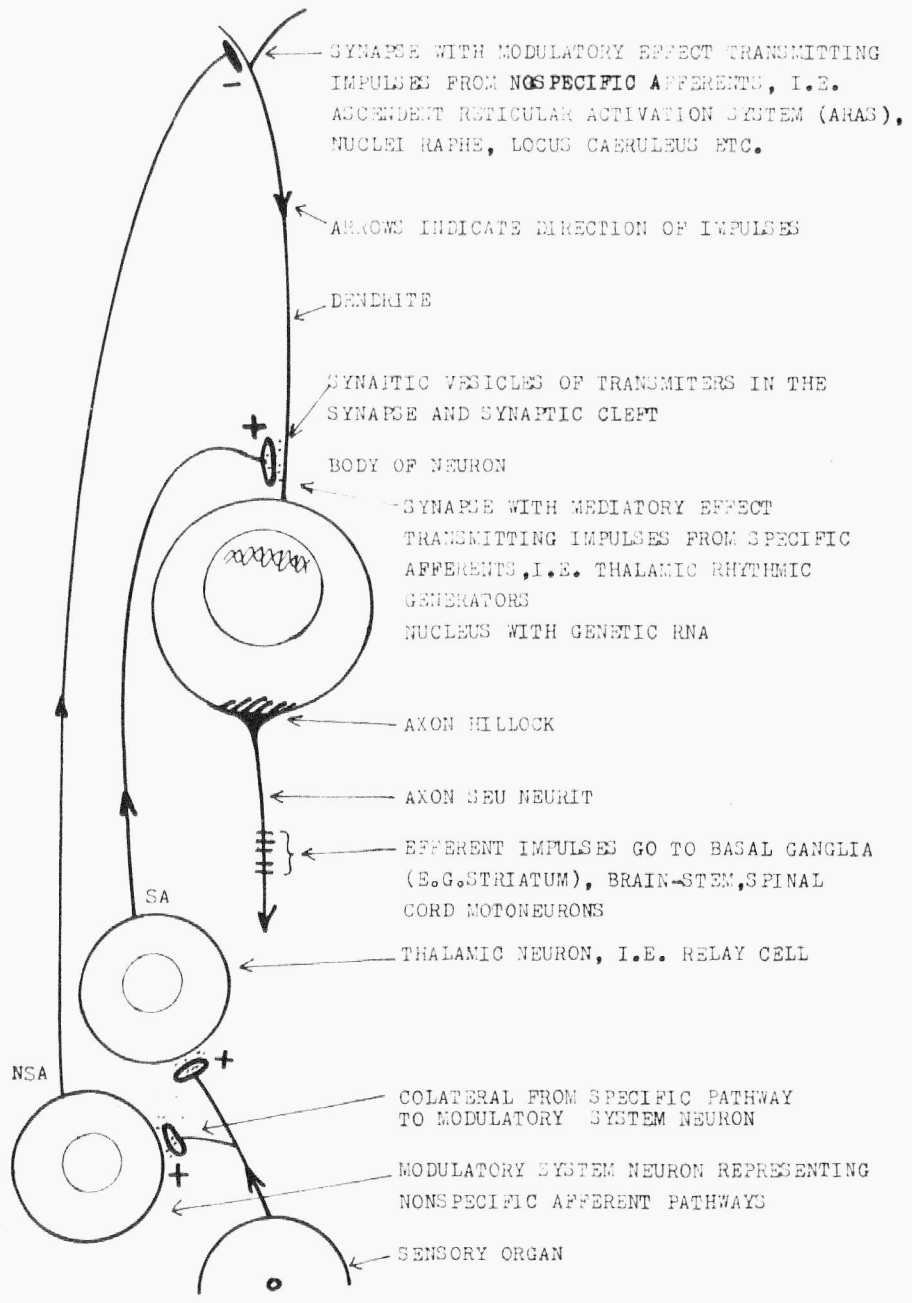


Fig. 7. Schematic view of cortical pyramidal neurons with their organeles (synapses, dendrites, body of neuron, axon hillock, and neurit). Connexion between pyramidal neuron and specific projection afferent pathways (SA) and nonspecific projection afferent pathways (NSA) from the brainstem neuron.

or the shape of an object just seen, etc. Verzeano (1977) proved this for the frequency of flashes. In epileptic activity, the neurons involved produce primitive firing patterns, and as this precludes any meaningful interneuronal communication, there is loss of consciousness, reactivity, and data recollection from and storage to memory (Faber 1975, 1978).

4. The above described regimen is controlled by brain stem modulation nuclei, whose impulses reach the target structures (cortex), thus introducing a definite programme responsible for functional changes in the neuronal networks within permissible limits. This can best be shown on an example of altered perception and mentation. During vigilance, the sound of a locomotive whistle sugests that the locomotive is near, during NONREM the same sound evokes a dream about a railway station, and — during REM sleep — say, a nightmare with the hooting of an owl. In other words, on and the same neuronal network in the temporal lobe has changed under the effect of formators so much as to interpret one and the same environmental sound in entirely different ways.

4. Anatomy and function of different parts of the brain

4.A. Thalamus — generator of signals. From the peripheral sensory organs, impulses pass along the previous interstations on to the thalamus. This is the last point before reaching the cortex that the impulses are relayed from neuron to neuron. The thalamus contains many nuclei, for each sensory quality one or a group of nuclei, i.e., for the sight, hearing, skin sensitivity, and so on. Each nucleus is divided into small groups of neurons called rhytmic thalamic generators. Each group receives impulses from a small limited cluster of sensory cells concentrated, say, around one cilium, or from a hundred rods in the retina.

The thalamus of the cat contains some 25 thousand such generators, the human thalamus about one million. It is from there that impulses, once "re-encoded", travel on to the cortex. Thalamic "relay cells" do not relay impulses immediately but rather after previous accumulation of impulses and in accordance with their momentary threshold. This threshold is raised, e.g., during sleep. A thalamic cell is influenced also by non-specific afferentation from brain stem nuclei and by two inhibitory cells nearby in what are typical examples of feed-back and feed-forward relationships. One inhibitory interneuron receives collateral information from the relay cell itself to exert inhibition on that relay cell *(Fig. 8)*. The other inhibitory interneuron receives collateral information from a specific afferentation, and inhibits the first inhibitory interneuron.

The result of this activity is as follows: once it has received a supracritical quantity of impulses, the relay cell sends an impulse to the cortex. By way of a collateral, however, it will excite the first inhibitory interneuron which causes hyperpolarization of the membrane of the relay cell, thus inhibiting it. This process can be blocked experimentally, e.g., with bicucoline or penicillin, in which case there is no inhibition, the relay cell is not inhibited, and every impulse will provoke a burst of impulses, which may result in epileptic activity. To avoid too much inhibition by the first inhibitory interneuron and to allow more impulses reaching the cortex while sensory organs happen to be supplying too much information from outside, the first inhibitory interneuron is inhibited by the other inhibitory interneuron in a feed-forward fashion (Andersson and Holmgren 1975).

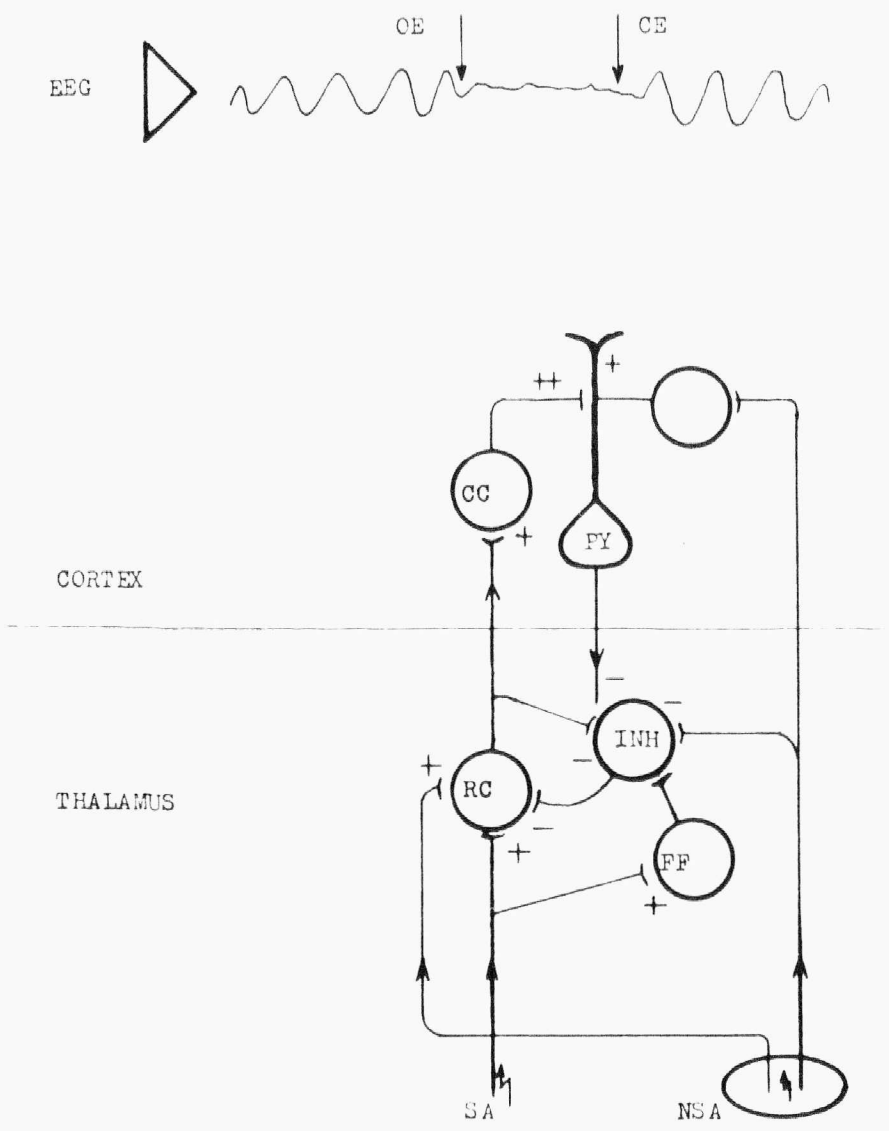


Fig. 8. Rhytmic generator in thalamus and its projection to cortical pyramidal cell. Typical vigilance EEG curve at the top is blocked during open eyes (OE) and unblocked after closing eyes (CE). PY = pyramidal cell, CC = interneuron with cartridge synapses, RC = relay cell, INH = inhibitory interneuron, FF = interneuron with feed forward inhibitory function, SA = specific afferentation, NSA = nonspecific afferentation.)

Once the inhibition is over, the neuron spontaneously aims at depolarization or postanodal exaltation, i.e., at further discharging which is enhanced by excitation impulses coming in from any of the sensory organs. The discharge is followed by another spell of inhibition, thus starting a cycle of discharges and void intervals. This rtythm may assume different frequencies, about 5 up to 25 Hz, i.e., in the theta, alpha up to beta bands.

Thus, due to the cumulation of impulses, the thalamic relay cell exhibits short bursts of impulses reminiscent of clock pulses. The thalamic generators are

plentifully interconnected, hence the impulses are propagated to other areas of the thalamus and to the whole of the remaining part of the cortex. This keeps the whole cortex informed of all sensory qualities.

In reality, the situation is more complex as all the above listed neurons of the thalamus receive excitatory and inhibitory impulses also from non-specific regions of the brain stem, i.e., from the modulation systems for vigilance and sleep, from the cortex and from the motor systems, i.e., from the cerebellum and basal ganglia. The thalamus is a very important relay station (body periphery — cortex) but also a significant integrating unit. Hence, as already mentioned, its name — the gateway of consciousness. According to Penfield (1969), movement is initiated in the thalamus, too.

4.B. Cortex — complex. Impulses evoked from the thalamic relay cells and their groups — rhythmic generators — reach the cortex, in particular layer 4, where they stimulate excitatory interneurons with large synapses, the so-called cartridges with their ability to amplify the impulses arriving there, and to stimulate powerfully the principal dendritic trunk of pyramidal cells of layers 3 and 5. If the stimulation is above the threshold, the pyramidal cells will send their impulses from within the cortex, in particular to the basal ganglia, the brain stem motor nuclei, and the spinal motor cells. Pyramidal neurons of the 3rd layer send typical subcortical association fibres to the remote areas of the cortex, or they run through the corpus callosum to the opposite hemisphere, and their fibres terminate again in cortical layers 2 and 3. The pyramidal neurons of layers 3 and 5 also have inhibitory neurons of their own similar to the thalamic relay cells, and these inhibit them by way of their feed-back mechanism. *(Fig. 9).*

More collaterals arising from the pyramidal cells project to the excitatory interneurons with a "bunch" of dendrites. Hence there French description "double bouquet dendritique", whose neurites stimulate profusely the neighbouring neurons and, consequently, also the dendrites of pyramidal cells. In this way, a "self-sustained loop" is developed. This activity concerns neurons of one column, i.e., about ten thousand neurons organized in a vertical little column running all through the cortex layer. This mode of impulse propagation would mean a risk of the whole

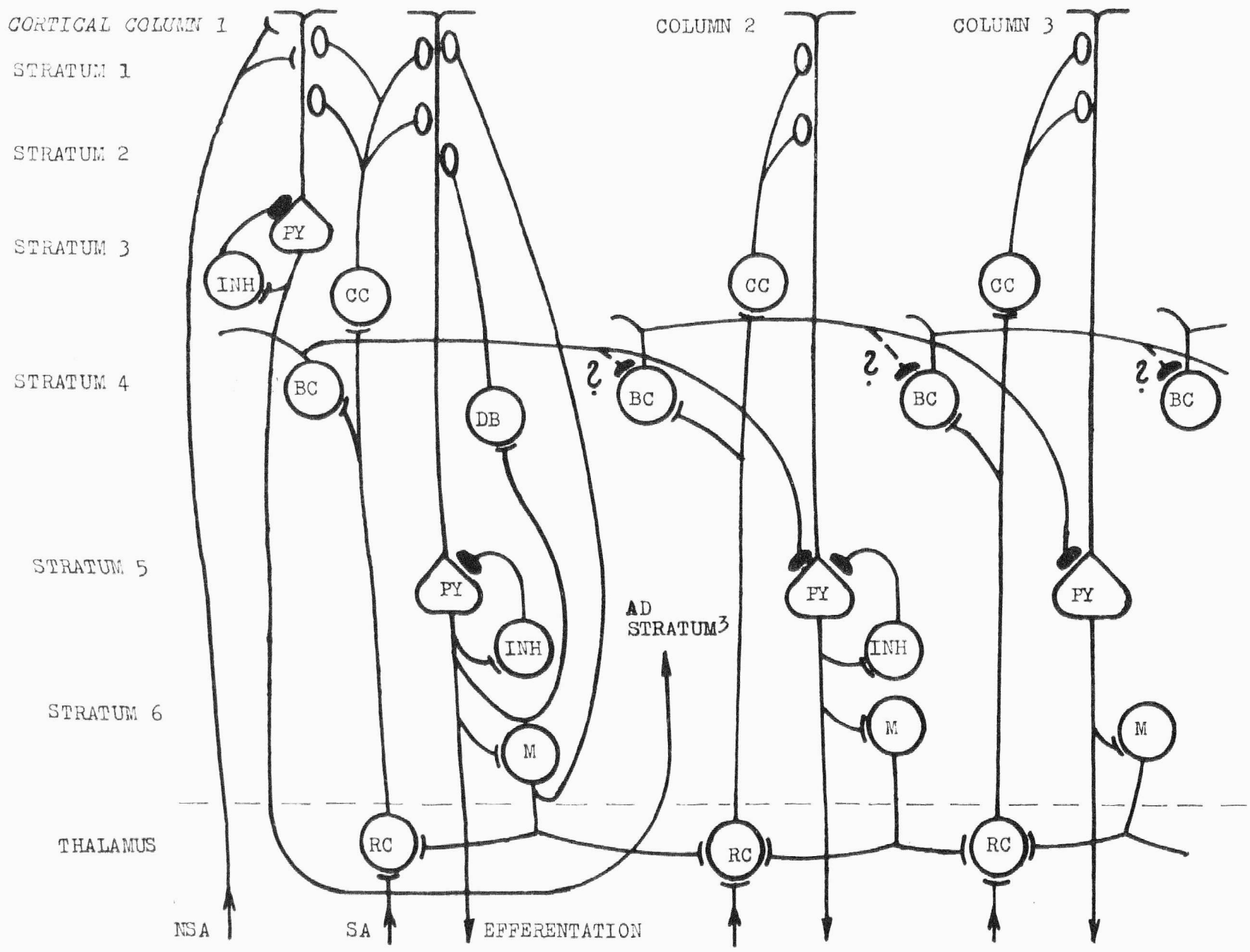


Fig. 9. Structures of the cortex and thalamus. The six-layer organization is discernible in the cortex. Layers 1 and 2 receive NSA from the brainstem centres for vigilance and sleep, layers 2 and 3 association and commissural fibres from other parts of the ipsilateral and contralateral hemisphere, layer 4 -specific afferentation (SA) fibres from, say, the eye by way of thalamic relay cells (RC). Layer 5 sends out efferents of the large pyramidal cells (PY) to spinal cord as the corticospinal tract. Layer 6 projects neurits to the thalamus. (BC = basket cell, DB = "double bouquet dendritique" neurons, M = Martinotti's neurons. The question-marked dashed line stands for inhibition of inhibition as we proposed and simulated it (Faber and Weinberger 1988). (Other abbreviations the same as in Fig. 8.)

cortical system getting off balance under the effect of the preponderant positive feed-back, and developing into an epileptic state. For that reason, there is more inhibition in the shape of basket cells. An axon projecting from the thalamus stimulates, in part, interneurons with "cartridge" synapses, in part, with its collateral, those basket cells which fibres pass through the cortex horizontally inhibiting pyramidal cells of layers 3 and 5 in the neighbouring columns. In this way, inhibited surroundings come into existence around the excited column.

The ten thousand neurons of the column include about 800 pyramidal cells, and out of these, some 100 are needed to make the excitation rise above the threshold and initiate muscular movement (Eccles 1973). Several columns together make up a functional whole called a hypercolumn. Within this higher entity the columns alternate, e.g., in motor areas for the extensors and flexors of the extremities. Their activity can be visualized as follows: *stretching the arm forward will activate all the odd-numbered columns — 1, 3, 5, 7, etc. — which innervate extensor muscles, while the even-numbered columns are inhibited by the basket cells. As the object is grasped by the hand, the even-numbered columns for the flexors of the arm* and the hand become innervated while the odd-numbered columns are now inhibited. One of the results of our modelling was the idea to introduce an inhibition of inhibition. After that, the model of thalamo-cortical reverberation proved to operate with far more reliability. The point was that the basket cell inhibited not only the neighbouring pyramidal cells, i.e., cells of column 2, but also the neighbouring basket cell, thus disinhibiting other pyramidal cells in the 3rd column (Faber and Weinberger 1988).

The hypercolumn in the visual cortex has a complex function to perform. Every column perceives lines slanting at a different angle. Their synthesis then permits, e.g., the reading of letters (Hubbel and Wiesel 1962). The synthesizing neurons exhibit a hierarchical arrangement: with simple neurons registering differently oriented lines, from where fibres project to complex neurons which register lines moving at a variable angle. And last, hypercomplex neurons receive fibres from complex ones to register, for example, angle-defined areas. Up to this point, experimental evidence is available. From there on the neuronal hierarchization can be visualized as progressing further on so that we can imagine hyper-hyper-complex neurons designed to identify letters, houses, persons, and so on *(Fig. 10)*. Taylor (1990) has devised a model of retinal and cortical processing.

The human cortex is a wonderful structure containing about 20,000,000,000 neurons aranged in a thin

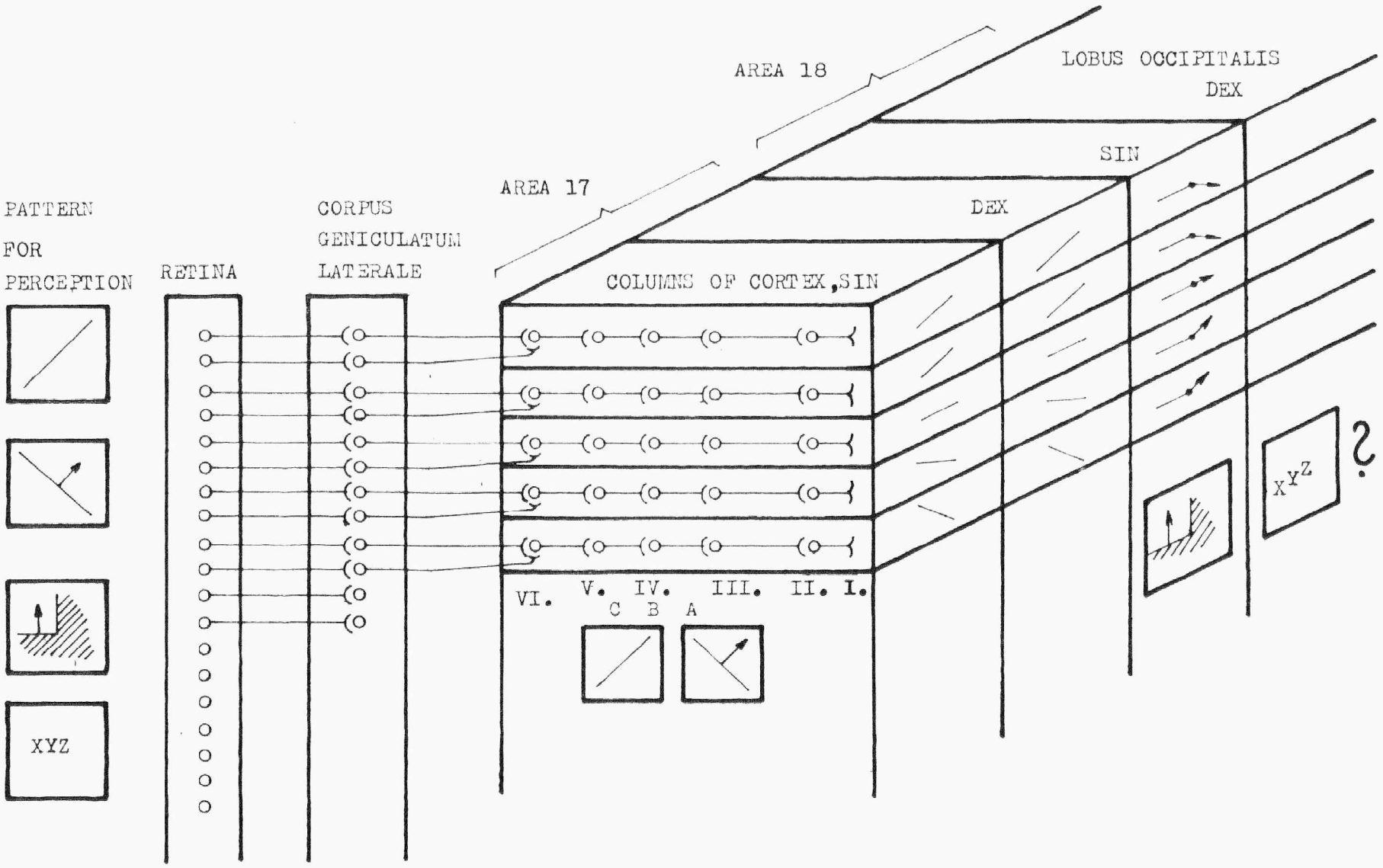


Fig. 10. Structure and functions of the optical cortex. Spaceorientated lines or sectors of surface are projected onto the retina, transmitted via the corpus geniculatum laterale to the cortex. This is where a simple analysis of the lines, angles and edges is made in the deeper strata of layer 4 (c) and a more complex analysis in the shallower strata of layer 4 (b, a). Series of neighbouring columns perceive the different orientation of lines. The adjacent columns running parallel perceive the same from the contralateral eye. Columns of similar function localized in series or parallel to each other constitute a hypercolumn.

layer of 2—5 mm in thickness and nearly a quarter of a square meter in size. One cannot help regarding the cortex as a kind of mega-chip with its individual component parts extremely ingeniously integrated. The cortex itself is non-independent, but it has attachments such as the signal generator and the data bus for addresses and concrete data, which is taken care of by the thalamocortical system. Also involved in the cortex are modulation brain stem systems which represent instructional data and have a data bus of their own (fasciculus longitudinalis telencephali = medial forebrain bundle). Another addressing attachment can be seen in the limbic system (LS). The LS works at the level of a higher programming language. Without it, the function of the thalamocortical circuit is very primitive and possible only at "assembler" level, though quite confusing in terms of intelligent communication. For example, a patient with his limbic system pathologically involved is capable of speech, but unable to *recall things from memory, given to confabulation, and emotionally superficial. But he is still conscious. A patient with the thalamocortical system pathologically involved is unconscious.* The last known address and data bus is the subcortical associative system of the so called "U" fibres.

If we compare the cortex to a chip, then the layers of the cortex have the following functions:
layers 1 and 2 represent the input for the bus of instructions from the brain stem modulation systems, i.e. from the formators;
layers 2 and 3 are a data bus and amplifier and bus of addresses from the remote columns of the cortex of the same layer (associative and commissural connections) as carriers of what are already partially processed data from other columns;
layer 4 is the input of concrete data from the thalamus as carriers of concrete data from sensory organs as well as thalamic clock impulses, but also as carriers of processed data from other columns via the thalamus;
layer 5 is the output of data from the cortical column; it is an executive function, which means that these impulses can pass through other nuclei to muscles or to basal ganglia where they pre-programme and perhaps even remember motor patterns, "know how";
layer 6 projects neurites to the thalamus, thus closing the thalamo-cortico-thalamic circuit. In this way, data are processed between the columns via the thalamus.

Connetions from the 6th layer represent communication between the cortex-complex and the thalamus-signal generator. They are the largest data bus in the brain, transmitting data between the different parts of the chip, i.e., between the columns, and taking care of impulse divergence, i.e., propagating any information throughout the brain in a few tens of milliseconds independently of the specific sensory organ (eye, ear, skin). The primary cortical projection areas for each sense are small. They receive information from the senses the earliest, within 10 to 30 msec, most of the rest of the brain, the motor and association areas soon afterwards, in about 100 msec. Connection with the LS, especially the hippocampus, is via the thalamus, hypothalamus and mesencephalic tegmentum, which represents potentially fast recollection from memory. The results of this activity are recognition and identification of that which is seen, heard or felt, associative thinking, and mode of response to a situation analyzed. The speed of the response to a stimulus is determined by the "intelligence" of the system, age, experience, "complexity of the scene", etc., and, as a rule, is anything between 150 to 500 msec. This amounts to about 2—5 reverberations, i.e., runs between the thalamus and the cortex.

4.C. Projective and associative pathways — data, address and instruction buses. We encounter about four types of buses. The LS effect on TCRC was already mentioned before. The main subject is the relationship between reason and emotion.

Another instruction bus is represented by brain stem modulation designed to regulate the main states of consciousness, e.g., vigilance now, sleep at some other time. Similarly to the first one, this too is to a prominent degree determined genetically.

The third data bus is determined anatomically by the subcortical intercolumnar associative connection linking layers 2 and 3 of different columns, i.e., fibrae ercuatae, fibrae commisurales. This is a sort of horizontal organization of connections designed to process data at a higher level which could be defined in terms of the theories of chaotic dynamics and cognitive channels with the aid of attractor and fractal functions (Nicolis 1987). These are physiological functions and anatomical structures developed under the effect of external factors, i.e., programmed by education and upbringing. Psychologically, they relate to abstract thought.

The fourth data bus, consisting of thalamo-cortico-thalamic pathways, a kind of vertical organization of data apparently designed for primary data analysis, represents real thinking. This structure is, to a large extent, determined congenitally.

With some exaggeration we might say that in the absence of the thalamus but in the presence of a sound cortex we would be capable of abstract thought and data synthesis. And vice versa, given a good function of the thalamus and cortex but in the absence of subcortical associative fibres there would be problems with synthesis but there would be a relatively good primitive analysis of data. The latter alternative has its clinical version — periventricular atrophy in dementia.

In a normal brain, the last two data and address buses appear to take turns in their activities, keeping in regular time. Stroke 1 is marked by primary information reaching the cortex from the vertical thalamic data bus. The information is processed as the impulses reverberate between the thalamus and the cortex, probably giving rise to primary primitive images of association, e.g., we recognize seeing a man wearing a hat. From the primary sensory area the impulses travel into the primary and secondary associative

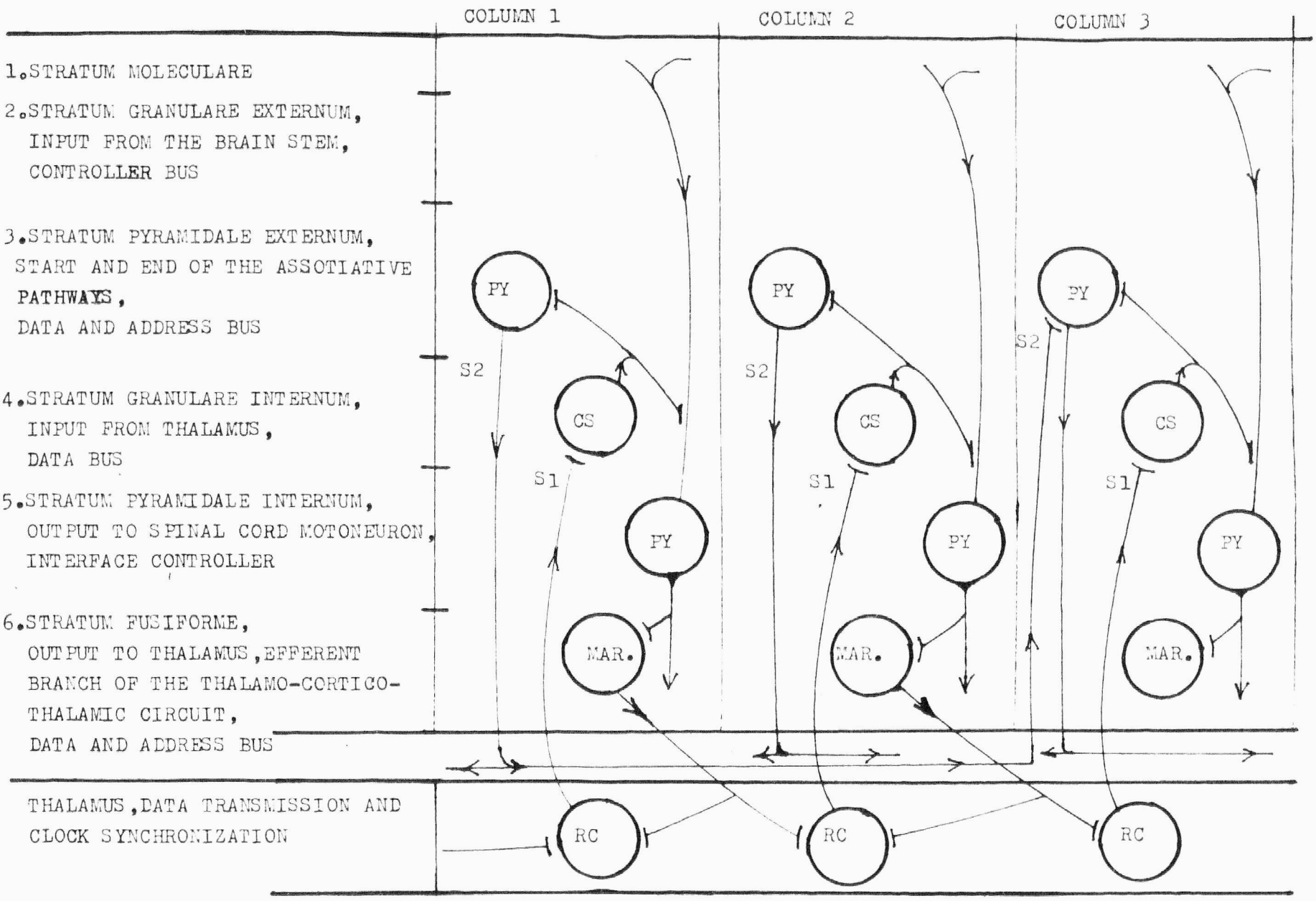


Fig. 11. Cortex as a two-stroke analyzer. Columns and strata keep the time: 1. stroke one is a thalamocortical impulse afferentation, intercolumnal connexion via thalamus, 2. stroke two is a subcortical impulse association between one and another stratum pyramidale externum, intercolumnal connexion via fibrae arcuatae, assotiativae et commisurales. (S1 — stroke one, S2 — stroke two, MAR — Martinotti's neurons, CS — neurons with cartridge synapses). (Abbreviations the same as in Fig. 8, 9).

fields and, from there, mainly into the hippocampus to look up earlier memory traces for comparison. Following stimulation of layers 2 and 3 in the cortex, stroke 2 is started automatically, activating the horizontal data bus. Partially processed data from a particular group of columns then pass along subcortical projections to more columns and areas for further more detailed and more general processing. Thus, for instance, we recognize the hatted man as an acquaintance. Anatomically speaking, the impulse passes from primary sensory area 17 with its simple neurons towards the complex and hypercomplex neurons of areas 18 and 19 of the visual cortex. *(Fig. 11).*

Stroke 1 has a cycle-like alpha activity, i.e., in the region of 10 Hz. Stroke 2 has the delta cycle, i.e., 1—4 Hz. The cortex in itself without the thalamus is marked by delta as the dominant activity (Stein 1965, Rappelsberger et al., 1981). The results of strokes 1 and 2 appear to be synthesized in tertiary areas, i.e., for the visual cortex — in area 19, for the acoustic cortex — in area 20, for the tactile cortex — in area 7. To corroborate this speculation, the technique of evoked, especially cognitive (P 300), potentials could be used to advantage (Sutton et al. 1965, Pollich et al. 1986). These are evoked potentials registered similarly as EEG from the cranial surface but visualized only after many (about 100) repeated stimuli have been averaged. This is a positive wave accompanying a conditioned reflex, with a latency from 250 up to 450 msec after the stimulus. The specifications of this wave are stimulus-dependent. Its amplitude and latency show the degree of difficulty of recognition of two different stimuli. The amplitude is directly proportional to the significance of the stimulation and to the importance of the motivation. It is low in states of depression. The longer the latency of the wave, the more difficult it is to discriminate between two stimuli. The latency grows in cases of dementia. A comparison of normal persons' latencies registered at different points of the cranial surface would permit taking a more qualified view of the existence and function of the "data buses". P 300 monitoring is already now made use of in clinical practice.

As the above facts suggest, the brain cannot be viewed as a single-chip computer, though there are objections to the idea of the cortex being one chip. The cortex is composed of neurons, and these elements cannot be likened to simple flip-flop circuits or arithmetic-logical units realized by means of transistors. Every neuron is in itself rather a small hybrid

computer with a digital-analog input and analog-digital output in the axon hillock. The neuron itself appears to operate mostly in the analog mode and to have its memory in the form of protein structures which regulate the input and output thresholds.

Interneuronal connection is effected by means of dendrites and neurites or axons terminating in synapses. The surface membranes of those fibres and the neuron bodies act as conductors for physical signals (electrochemical processes) or impulses which carry information and are frequency-modulated on the axons, and analog-modulated on the other parts of the neuron. On entering the neuron, bursts of impulses are turned into continual analog-changes in the membrane potential. An excitatory synapse will cause depolarization which, once it has reached the critical threshold of some 40 mV, will result in precipitate depolarization and transpolarization of the membrane, whereupon from the axon hillock where the axon begins the neuron will send an impulse to other cells in what is called analog-digital conversion. An inhibitory synapse follows a precisely opposite course as its impulses lead to membrane hyperpolarization (of up to 90 mV or more) with the neuron being inhibited, i.e., remote from a discharge. A neuron has many dendrites (inputs) and one neurite or axon (output). (Kandel and Schwartz 1985).

Far from being stationary, the threshold of neuronal inputs and output keep changing under the effect of synaptic potentials. In this respect, modulation synapses appear to have a greater effect than mediator synapses. In this way, brain stem modulation centres (formators) alter neuronal activity and, thereby, also the programme of neuronal circuits and networks. This causes changes in vigilance, sleep and, pathologically, also in paroxysm or psychotic raptus.

It appears then that the cortex can be likened to a "non-von Neumann" system, i.e., rather to a transputer. The neuron is more than just a flip-flop circuit, it is a computer with a memory of its own designed to process information independently and pass it on or withold it, in other words, "to make ad hoc decisions of its own". There are many neurons dealing with similar jobs, a phenomenon well known from the visual cortex (Kuffner and Nicholls 1976, Hubel and Wiesel 1974) and referred to as redundance in the nervous system. Among other things, this means the parallel involvement of very similar, though not quite identical, microcomputer units and a better analysis resulting in an isomorphous representation of reality. It is quite probable that this is only apparent rather than real redundancy. The drop-out of some fine structures need not be clinically discernible, albeit identifiable through very detailed psychophysiological testing.

Compared with a technically devised transputer, the brain has the additional complexity of four types of regulators — buses — taking part in the control of the complex-cortex: thalamo-cortical, subcortical, limbic and brain-stem regulators.

4.D. Epileptic focus (EP)

An epileptic focus is a group of neurons featuring special, abnormal properties. As a rule, an EP is localized in the neo- or archi-cortex, and arises due to any pathological process, injury, asphyxia, inflammation, vascular disorder, etc. The pyramidal neurons in the focus usually exhibit lesions of the dendritic trunk and synapses, especially axosomatic inhibitory synapses. Frequently, there are disorders of the neuroglial cells, i.e. satellite, supporting cells, i.e., structures designed to take care of neuronal metabolism (Fischer et al. 1968, Ward 1961). As a result, there is a damaged neuron in a parabiotic state which is insufficiently inhibited either by mediators or modulators, and, consequently, mostly in a state of moderate depolarization and with a tendency to keep firing. Hence the EP behaves like an aberrant formator competing with physiological formators for influence over the cortex.

The discharges of such a neuron differ from those of a normal cell. They are faster and exhibit a different firing pattern with short inter-impulse intervals, shorter than 5 3 msec, and with primitive rhythmicity. A firing pattern like that is a threat to other neurons as it causes rapid depolarization of other connected neurons and their potential lapse into an epileptic regimen. A discharge of this kind then propagates at a geometric rate. As Calvin (1972) found out, a mere one per cent of impulses of this firing pattern on entering a normal neuron will do for the normal neuron to become epileptic itself. In the brain there is a constant danger of the increase of damaged neurons threatening to drive the whole brain into pointless primitive firing and, consequently, into an epileptic seizure. Hence the presence of an effective system of defence. There are not only inhibitory interneurons operating as "satellites" to every larger neuron but also quite large structures such as the caudato-thalamic, rubro-thalamic or noradrenergic-locus caeruleus systems.

An epileptic focus contains about 10 % morphologically damaged neurons which keep firing solely in the epileptic mode, about 50 % normal neurons and 40 % neurons firing in the epileptic mode only occasionally (Lockard and Wyler 1979). Whether the focus will be only slightly or very intensive and whether their activity will actually lead to an attack depends on the 40 % facultatively firing neurons, i.e., on which side they will take. In EEG and stereo-EEG from implanted electrodes we register epileptic foci which, in clinical terms, are "silent" for longer periods of time in epileptics or permanently in psychotic patients (Faber and Vladyka 1984, Faber and Vladyka 1987), *(Fig. 12)*.

Under normal circumstances, the cortex is an anatomically and functionally accomplished integral whole interconnected by subcortical, thalamocortical, limbic and brain-stem pathways ("data buses"). An EP becomes emancipated, isolated, relatively autonomous and independent of the whole, uninhibited and uncontrolled either by the cortex or by other modulations. Moreover, it itself puts out many fast impulses

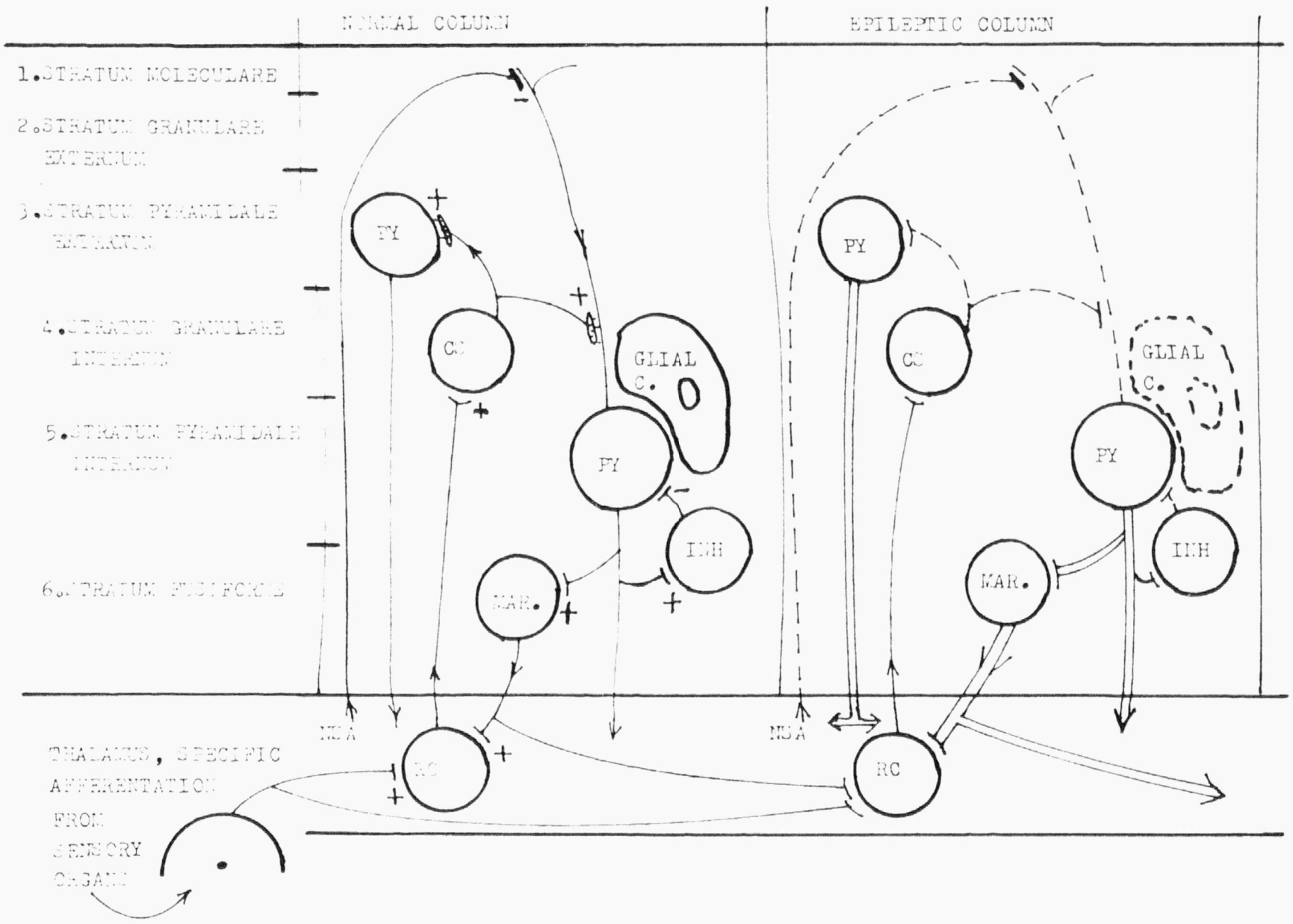


Fig. 12. Hypothetical structure of epileptic focus. The left column exhibits normal structure and functions, in the right column with glial cells (GLI-AL C.), dendrites and inhibitory synapses destroyed. Double lines of neurites mean augmented impulsation. Dashed lines of neurites and dendrites mean diminished impulsation.

whose primitive programming is readily acceptable to healthy neurons. An EP processes the information received inadequately, e.g., it is increasingly excitable during photostimulation (Panayiatopoulos et al. 1972).

Normal data and address buses also adopt this simple mode of work. As a result, more secondary and tertiary epileptic foci arise with the help of subcortical pathways. Thalamocortical pathways likewise adopt this mode of firing to pass it on to the whole cortex. The latest systems designed to inhibit this type of epileptogenesis are the brain-stem modulation systems, which are inhibitory by their very nature. If even those structures fail there is a manifest epileptic paroxysm. Specifically, the system of paradoxical sleep fails during epileptogenesis (Libet et al. 1977, Faber and Vladyka 1984). A similar mechanism underlies the development of psychoses. About one third of psychotics have an EP deep in their brain (Sem-Jacobsen et al. 1955, Trimble 1984, Heath 1975, Faber and Vladyka 1987) and also altered brain-stem modulations, which is manifested by lack of NONREM sleep in schizophrenia (Caldwell and Domino 1967) and by REM sleep changes in depressive syndromes (Kupfer et al. 1985).

Experimental models of "psychosogenesis" in animals (Stevens and Livermore 1978) show that the right kind of manipulation of brain-stem modulations can change the animal's "nature" and make its behaviour "funny". This can be attained by chronic stimulation (kindling) of mesolimbic structures leading to hyperdopaminergy and inhibition of the frontal lobe. As a result of this, certain symptoms of human schizophrenia appear, e.g., autism. Most drugs designed to control schizophrenia are of the antidopamine type, e.g., phenothiazines or butyrophenones.

4.E. Diencephalo-temporal circuit

The diencephalo-temporal or septo-hippocampal circuit (SH) is part of the limbic system. It represents structures phylogenetically older than the thalamocortical circuit and, in terms of function, it is dominant in reptiles, birds and, partially, in primitive mammals. The SH regulates emotions and behavioral automatisms and shares in data storage in memory. The hippocampus is regarded as a register of memory, i.e., determining when and where events happened, giving them the idex of time and space (O'Keefe and Nadel 1978). Pathological functions take the clinical form of temporal epilepsy or psychotic states (Servit 1983, 1987).

As already mentioned, the LS appears to be the site of interaction between impulses and incentives in the

motivation play. Objects of the external world have no absolute value in themselves; it is man that gives them this "value" by wanting or not wanting them. This interplay of "desire or rejection" is defined in terms of motivation, i.e., interaction of impulses and incentives or interplay of inborn and acquired tendencies. This "play about human happiness" appears to be co-determined by another factor, a set of general intellectual qualities, most of them congenital but developed by education and localized in the thalamocortical system. The general term for this factor is talent. Impulses, incentives and talent may sometimes be at variance. A harmonious personality is marked by those phenomena being well balanced.

Anatomically, the SH circuit includes most of the limbic structures: nuclei septi pellucidi in mammals, area adolfactoria in humans, nuclei interpeduncu-lates, nucleus habenulae and amygdala interconnected by the stria medullaris and stria terminalis. The SH further includes the hippocampus, corpora mamillaria, nucleus anterior thalami and gyrus cinguli; these structures are interconnected by the fornix and fasciola cinerea, and make up the Papez circuit (Papez 1937) and the Fisher-Curry circuit (Valzeli 1980, Maršala 1985). *(Fig. 13).*

The physiological EEG activity of the human SH system is probably very much like that in all primates, with fast theta and slow alpha (7—9 Hz). This activity is imparted by rhythmically firing neurons in the septal or adolfactory areas. This system secretes on the synapse acetylcholine as the mediator, which accounts for its description as acetylcholinergic. Its axons come by way of the fornix where there are also important fibres from the corpora mamillaria for memory and sexuality. Other projections to the hippocampus pass along the perforant path carrying information from the area entorhinalis which, in turn, receives information from the primary and secondary associative fields. These impulses are non rhythmical and carry information mediated from sensory organs.

The perforant path terminates in excitatory synapses on the apical dendrites of hippocampal pyramids and granular cells of the fascia dentata, which is part of the hippocampus. Fibres from the fornix touch with their excitatory synapses the trunk of dendrites and granular cells of the hippocampus. The axonal collaterals of the cornu Ammonis (CA) pyramids and granular cells of the fascia dentata project to the basket cells and from there back to the same pyramidal cells. This is a well known negative feedback — recurrent collateral inhibition, the only inhibition in the hippocampus. Other excitatory connections include moss fibres from the granular cells and Schaffer's collaterals, i.e., fibres projecting from CA3 into CA1 in the hippocampus (Eccles 1973). Schaffer's collaterals have a powerful excitatory effect, the kind of kindling going on here is very effective, and epileptogenesis takes a rapid course (Wadman et al. 1983). Due to the fact that in the hippocampus there is a predominance of excitatory over inhibitory synapses at a ratio of 4:1,

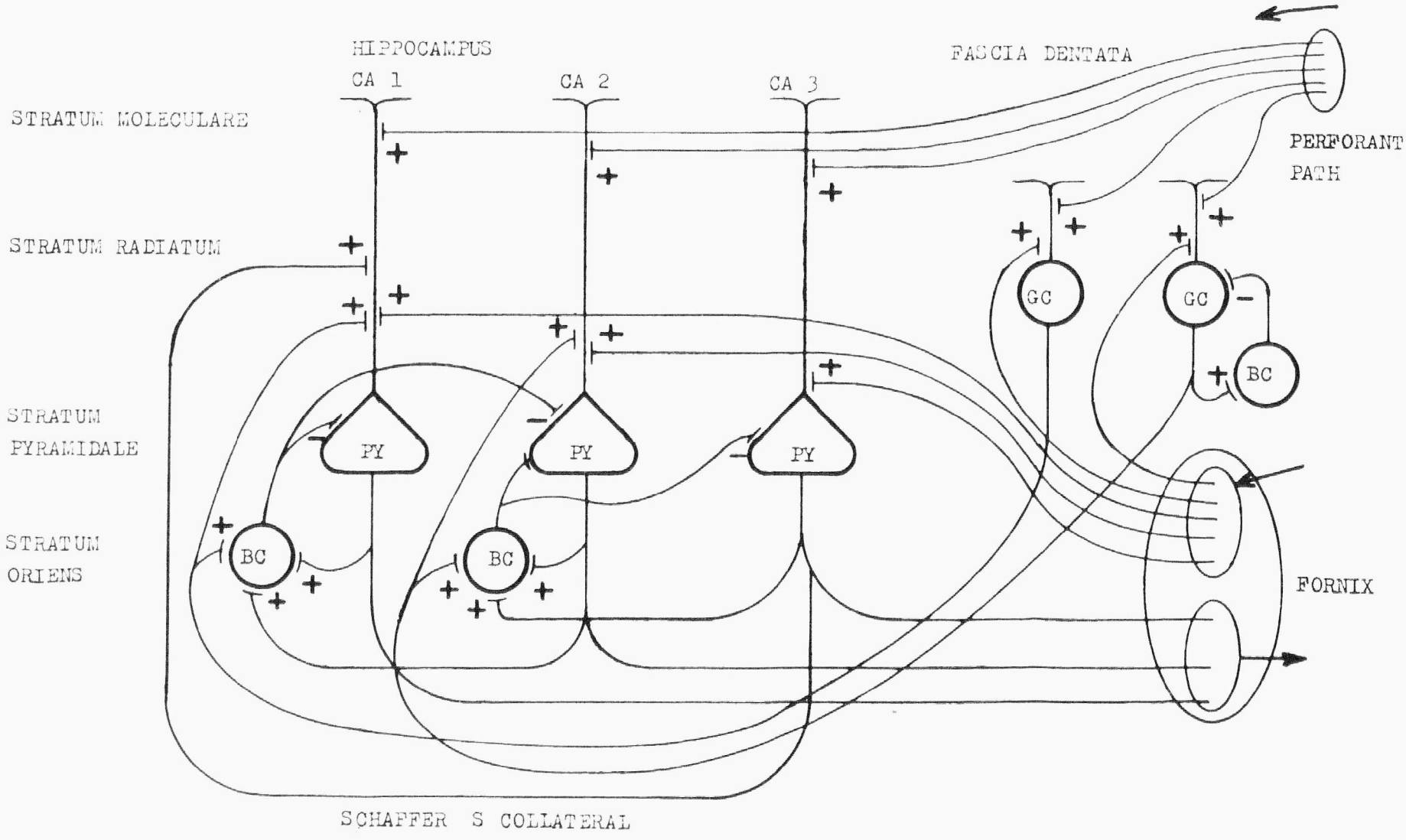


Fig. 13. Schema of hippocampal structure showing its three-layer organization. Pyramidal neurons exhibit four excitatory and one inhibitory afferentations. CA = Cornu Amonis, GC = granular cell. Other abbreviations the same as in Fig. 8, 9.

this structure is noted for great excitability, i.e., a low paroxysmal threshold. This was repeatedly shown in classical experiments with strychninization and, more recently, with chronic electric stimulation inducing a kindling effect (Goddard and Douglas 1975).

The LS includes also a neocortical part such as the gyrus cinguli and fronto-orbital cortex. These are the sites for psychosurgical operations, as artificial lesions made there often bring considerable relief to psychotics suffering from anxiety, compulsive ideas and depression (Fulton 1951). Gray (quoted from Howard et al. 1982) described BIS and BAS, two systems closely related to anxiety and aggressivity. The BIS (behavioral inhibition system) arises from the following anatomical structures — gyri fronto-orbitales, nuclei septi and gyrus hippocampi and its purpose is to inhibit motoricity in conditioned reflexes and in complex behaviour. Hyperfunction of this system results in excessive inhibition felt subjectively as tension or even anxiety. Frontal lobotomy, i.e., separation of the frontal lobes from the rest of the brain, leads to suppression of anxiety but mostly also to personality devastation. Hence today's preference for but minor lesions made, as a rule, in the innominate zone, anterior capsule, the anterior part of the gyrus cinguli. This helps to suppress anxiety while the patient's personality remains intact (Laitinen 1974). *(Fig. 14).*

BAS stands for behavioral approach system. Anatomically, it comprises the lateral hypothalamus, lateral septal nuclei and the fasciculus longitudinalis telencephali or medial forebrain bundle, and also part of the amygdala.In terms of function, this structure underpins behaviour. Hyperfunction there leads to impulsive or even aggessive behaviour. Hence, psychosurgery designed to suppress aggressivity has its targets in the lateral and dorsal hypothalamus and in the amygdala.

4.F. On the one hand, nuclei of the brain stem exhibit an inconspicuous anatomy with simple clusters of neurons; on the other hand, not enough is as yet known about the interneuronal connections within the nuclei. In contrast, rather a great deal is known about the internuclear connections. These constitute a very complex network and permit the "formators" to take turns in their activities (Hobson 1977, Petrovický 1981).

5. Sensorium

Let us say that the term sensorium denotes consciousness (das Bewusstsein, conscience, soznaniye). From the psychophysiological point of view, we can see it as a phenomenon of three qualities, namely,

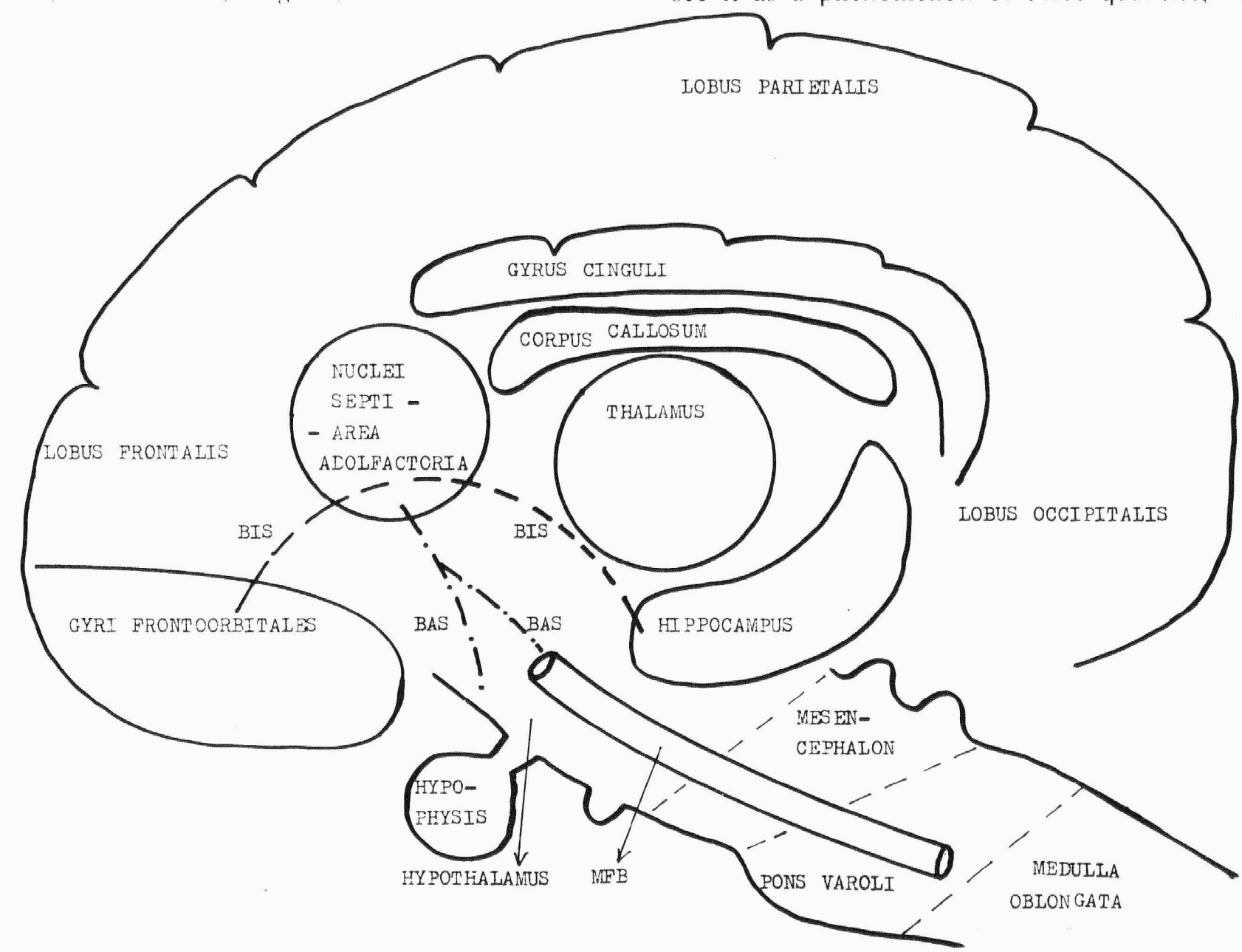


Fig. 14. Gray's BIS (behavioral inhibition) and BAS (behavioral approach) systems, two circuits, each combining emotional and psychomotor activities. (MFB = medial forebrain bundle, i.e. fasciculus longitudinalis telencephali, pathways for NSA.)

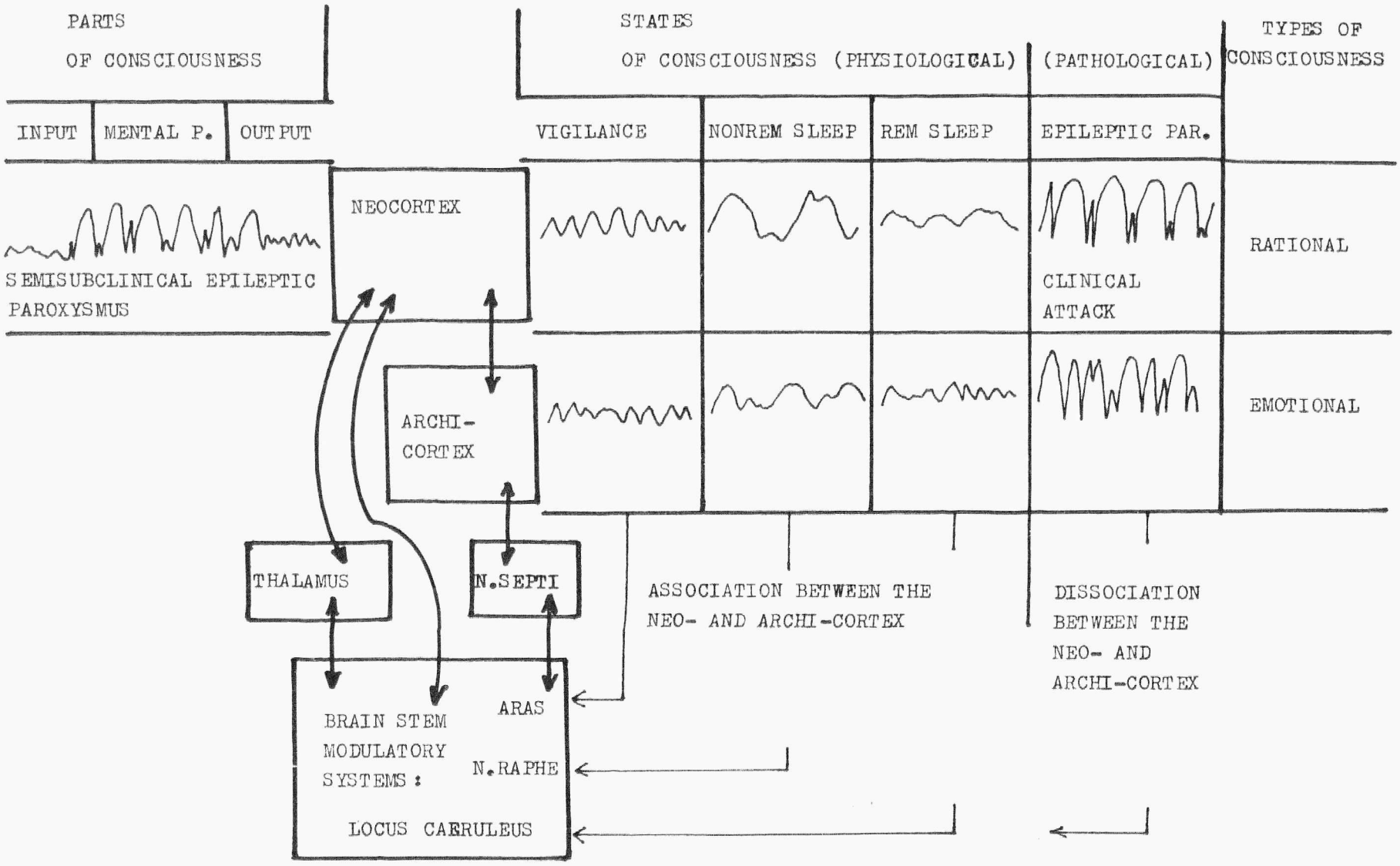


Fig. 15. Attempted interpretation of the complex notion consciousnessensorium, according to electrophysiological findings. Parts of consciousness can be classified as input: sensory function (seeing, hearing), mental processing: (MENTAL P.) (gnostic function, abstract thinking), output: psychomotor activity (behaviour, speech). States of consciousness are vigilance, NONREM sleep and REM sleep, pathological states are e.g. epileptic paroxysms or psychotic raptures. Types (or kindes) of consciousness — rational, emotional. Left above: EEG curve of semisubclinical epileptic paroxysm, the higher mental activity disappears, but simple reactivity to external stimulation remains. Right above: EPILEPTIC PAR., CLINICAL ATTACK, i.e. loss of consciousness.

quality of state, quality of type, and quality of parts of consciousness *(Fig. 15).*

5.A. States of consciousness — vigilance, synchronous sleep and paradoxical sleep — are so different from one another that they can be viewed as mutually independent states. They are inborn phenomena developed in the 7th month of foetal life, i.e., well within pregnancy, and defined by different regimens of the brain. Each exhibits a distinct metabolic and EEG activity. They are rhytmically repeated and take turns throughout each individual's lifetime. During wakefulness, the organism receives information from the environment and energy from food, during sleep, the process of learning continues with information being stored in memory, selected and abstracted. These are states of the lst order, and they are determined solely genetically. They follow a pattern of periodic activity controlled from specific centres in the brain stem as listed under 3.D. The archicortex and the neocortex operate somewhat differently from each other, but their activities are co-ordinated by means of brainstem modulation into mutual association. Disordered co-ordination results in dissociation and in a pathological state. For example, hypofunction of the neocortex may produce anxiety or hysterioform behaviour very much like in archicortical hyperfunction. Conversely, hypofunction of the archicortex may give rise to depersonalization and other symptoms of psychasthenia. Marked and prolonged insufficiency of the modulation systems produces deeper-seated disorders. For instance, NONREM sleep deficiency may lead to the schizophrenic syndrome. REM sleep insufficiency may result in epileptic attacks, and REM disorganization may provoke depression.

States of consciousness of the 2nd order are developed postnatally. They include focused attention, orientational attention, conditioned reflex development, concrete thinking. There are also corresponding electrophysiological manifestations: focused attention has its EEG counterpart in synchronous alpha or theta; orientational attention — desynchronized activity and beta activity (Gestaut et al. 1957).

States of consciousness of the 3rd order are specifically human conditions. They are mechanisms of speech or phatic functions, speech understanding, speech production, reading, writing; in general — mnestic functions, i.e., the ability to associate words and phrases with their meanings, or thoughts with their expressions. Also included in this category is abstract thinking such as logical reasoning, calculation, etc. Duffy et al. (1981) and later a number of other authors, too, were able to prove that this activity is like-

wise EEG-detectable by means of mathematical-statistical data analysis.

5.B. The type of consciousness is another quality, the rational and emotional aspects of consciousness. As already mentioned, the rational functions are situated in the neocortex, emotional functions in the archicortex, in particular, the hippocampus. Electric stimulation of the neocortex or epileptic discharges in the same structure will excite or inhibit gnostic or phatic functions, evoking, e.g., visions of faces or causing "speech arrest". Electric stimulation or epileptic discharges in the hippocampus can often provoke anxiety or fear, rarely also voluptuous experiences (Faber and Vladyka 1984, 1987, 1988).

5.C. Parts of consciousness constitute the third quality. I. the qualitative aspect comprises (a) sensory perception (analysis of things seen or heard), (b) mentation; this is just another aspect of the functions listed under "states of consciousness of the 3rd order".Here they are viewed as a separate category potentially related to another separate category, e.g., to REM sleep. We can see then that mentation exists even during this type of sleep, albeit quite different from mentation in wakefulness. And last, (c) motor efferentation or action response to stimulation, behaviour. II. the quantitative aspect relates to the gradual loss of quality: lucid vigilance, somnolence, stupor, coma. The quantitative aspect is classified either as physiological, i.e., the simultaneous inhibition of all functions as we go to sleep and during sleep, or as pathological, i.e., slow and simultaneous extinction of psychic functions in the course of extracerebral coma such as hepatic, diabetic or uraemic coma. In an epileptic paroxysm, these functions usually come to an abrupt end, but in some absences or pseudoabsences (possibly also in their temporal equivalents) only some functions will become extinct while others remain relatively preserved such as, for example, discernible loss of phatic functions (speech disturbance) in the presence of preserved simple reactivity (Faber 1975).

Similar situations may arise in cases of narcolepsy where, during lapse into sleep, there may be a brief spell of paradoxical sleep in the form of hypnagogic hallucinations. Or in the course of waking up, there may be a discrepancy between vigilance, already functional, and muscular atonia persisting after paradoxical sleep with the patient perceiving all this as a disagreeable state of paralysis. Here we refer to the phenomenon of sleep paralysis which even perfectly healthy persons may experience once or twice in their lifetime. All the above listed qualities of sensorium have their electrophysiological, especially EEG, correlates.

References

[1] Andersen P., Andersson S. A.: Thalamic Origin of Cortical Rhythmic Activity. PP 2C-00-118. In: Handbook of Electroenceph. clin. Neurophysiol. Ed.: E. Creutzfeldt, Vol. 2, Part C, Elsevier, Amsterdam, 1974.

[2] Andersson S. A., Holmgren E.: Theoretical consideration on the synchronization of thalamo-cortical activity. In: Subcortical Mechanisms and Sensorimotor Activities. Ed.: T. L. Frigyesi, Huber, Bern, 1975, pp 229—250.

[3] Beneš J.: On neural networks. Kybernetika, 26, 1990, No. 3, 232—247.

[4] Caldwell D. F., Domino E. F.: EEG and eye movement patterns during sleep in chronic schizophrenic patients. Electroenceph. clin. Neurophysiol., 22, 1967, 414—420.

[5] Eccles J. C.: The understanding of the brain. McGraw-Hill Book Company, 1973, New York.

[6] Faber J., Vladyka V.: Epileptogenesis and "Psychosogenesis", Antithesis or Synthesis? Acta. Univ. Carol. Med. 33, no. 3/4, 1987, 245—312.

[7] Faber J., Weinberger J.: Thalamocortical Reverberation Circuit simulation using the Simula Language. Acta. Univ. Carol. Med. 34, No 3/4, 1988, 149—248.

[8] Farley B. G., Clark W. A.: Simulation of self organizing systems by digital computer. Trans. IRE, PGIT-4, 1954, 76—84.

[9] Goddard G. V., Douglas R. M.: Does the engram of kindling model of normal long term memory? In: Kindling, Ed.: J. A. Waada, Raven Press, New York, 1975, 385—394.

[10] Heath G. R.: Brain function and behavior. The Journal of Nervous and Mental Disease. 160, 1975, 159—175.

[11] Hobson J. A., McCarlev R. W.: Cortical Univ Activity in Sleep and Waking. Electroenceph. clin. Neurophysiol. 30, 1971, 97—112.

[12] Hubel D. H., Wiesel T. N.: Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. (London), 160, 1962, 106—154.

[13] Hydén H.: The learning brain during a life-cycle, some biochemical and psychological aspects. Acta neurol. scandinavica, Supplementum 80, 62, 1980, 9—27.

[14] Jouvet M., Delorme F.: Locus caeruleus et sommeil paradoxal. Soc. Biol., 159, 1965, 895—902.

[15] Kandel E. R., Schwartz J. H.: Principles of Neural Science. Amsterdam, Elsevier, 1985.

[16] Kupfer D. J., Spiker D. G., Coble P. A., Shaw D. H.: Electroencephalographic Sleep Recordings and Depression in the Elderly. J. Amer. Geriat. Soc., 25, 1978, 53—57.

[17] Lockard. J. S., Wyler A. R.: The Influence of Attending on Seizure Activity in Epileptic Monkeys. Epilepsia. 20, 1979, 157—168.

[18] Moruzzi G., Magoun H. W.: Brain stem reticular formation and activation of the EEG. Electroenceph. clin. Neurophysiol. 1, 1949, 455—473.

[19] Nicolis J. S.: Chaotic Dynamics Applied to Biological Information Processing. Akademie Verlag, Berlin, 1987.

[20] Rappelsberger P., Rockberger H., Petsche H.: About the intracortical genesis of spontaneous activity: EEG:histological correlations inthe visual cortex of rabbits. Electroenceph. clin. Neurophysiol., 51, 1981, 73 P.

[21] Sem-Jacobsen C. W., Petersen M. C., Jorge A. Lazarte, Dodge H. W. Jr., Holman C. B.: Electroencephalographic rhythms from the depths of frontal lobe in 60 psychotic pacients. Electroenceph.clin. Neurophysiol., 7, 1955, 193—210.

[22] Servít Z.: Epilepsy, Avicenum, SZN, Praha, 1983. (In Czech).

[23] Sjöström R.: Effects of psychotherapy in schizophrenia. Acta psychiatr. scandinavica. 71, 1985, 513—522.

[24] Sutton S., Braten M., Zubin J.: Evoked-Potential Correlates of Simulus Uncertainty. science, 150, 1965, 1187—1188.

[25] Taylor J. G.: Modelling Visual Processing. In: Neuronet '90. International symposium on neural networks and neural computing. Ed.: M. Novák. Czechoslovak Academy of Science, Charles University of Prague. Prague, 1990.

NEUROCOMPUTING AND CONSCIOUSNESS

D. L. Koruga)*

Abstract:

This article deals with the problem of interrelation between neurocomputing and consciousness. Neurocomputing is approached from the aspect of the space-time structures, while consciousness is perceived as a link between states of mind and images of these structures in the brain. This approach leads to a relativistic model of information theory, and opens up the possibilities of linking *information* with *mass* and *energy*.

By considering neurocomputing and consciousness, a new field of science emerges which can be named: informational physics. In the final discussion, one extra problem is considered: Can a machine, as a form of artificial life, posses consciousness?

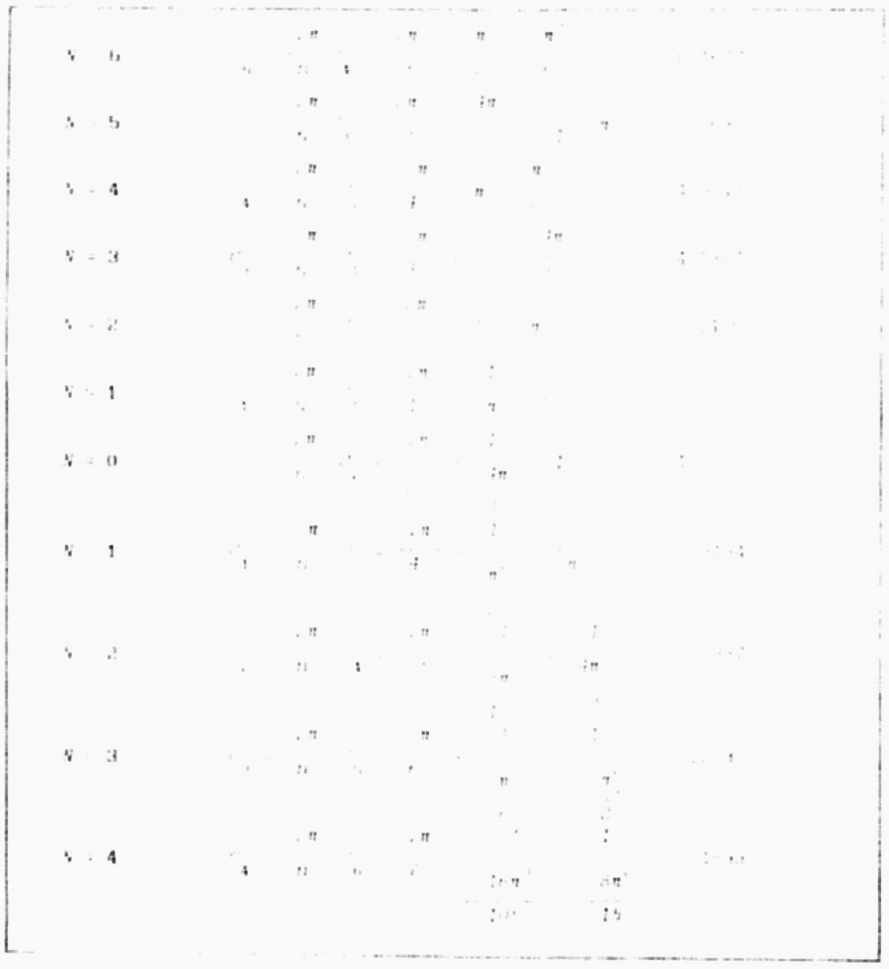
1. Introduction

From the point of view of importance, the question: What is conciousness?, comes after the questions: why are these essents rather than nothing? [1], and what is time? Consideration of the relationship between neurocompting and consciousness opens up new possibilities of closer answers to these questions.

In the scope of our recent research [2, 3] in the field of bioinformatics, we have shown that information processes as space-time patterns depend on the volume of the unit sphere in the N-dimensional space [4, 5]. The values of the unit spheres in N-dimensional space are given in *Table I*. In this table, appear the positive dimensions, the negative dimensions, and the dimension N = 0.

Considered from the informational aspect the optimal space is five-dimensional, and it appears as both positive and negative. Having in mind that a N = 0 space also exists, we will assume that the main informational state represents the unity of N = 5 and N = 0. $\kappa(5°) = 1$ where κ means the information code of dimension N = 5 and N = 0 (*Fig. 1a*). The unity sphere N = 0 contains the entity whose dimension is d = 3/2, and volume is $C_n = 1$, which is equivalent to the value of the space information code $\kappa(5°)$. In other words, there is a correspondence between $\kappa(5°)$ and C_n.

Table I: Value of the unit sphere for N = 6 to N = −4 Ref. 2, 3.



In order to establish a connection between the *informational approach*, the result of which is *Table I*, and the physical approach it is necessary to find a *physical entity* which, in five-dimensional space, has a dimension d = 3/2, and whose volume is $C_n = 1$.

It is well known from quantum field theory [6] that the dimension of mass, as a real physical entity, is calculated from the expression:

$$d_m = \frac{d}{2} - 1 \qquad (1)$$

where d is a dimensional space-time value. In the main information state $\kappa(5°)$ the mass dimension based on the expression (1) is $d_m = 3/2$. In other words, the real physical entity which contains information code is *mass* in the state $d(3/2, C_n)$.

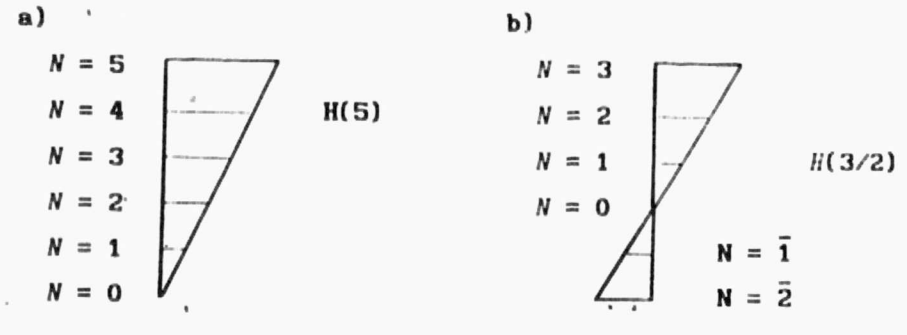


Fig. 1 Two solutions when N = 5: a) optimal and b) quasi optimal

*) Prof. Djuro L. Koruga,
Molecular Machines Research Center
Faculty of Machine Engineering
University of Belgrade
27. Mart 80, 110 00 Belgrade
Yugoslavia
E-mail: EKORUGA%UBBG@YUBGEF 51
Tel & Fax: +(3811) 320 207

From our results shown in *Table I*, it is implicated that N=0 as $d(3/2, C_o)$ may, from the aspect of information, pass into the new state $d(3/\bar{2})$, because $\mathbf{N}(3) \,.\, \mathbf{N}(\bar{2}) = \mathbf{N}(0)$. In this case we again have d = 5 in $N = \bar{2}$, but we also have d = 3 in N = 3 *(Fig. 1b)*.

We know that our awake consciousness is a result of *brain information processing* with interactions from the *real world*. We present every event of objective reality as:

$$x = x_1, \quad y = x_2, \quad z = x_3, \quad t = t \tag{2}$$

and this is our four-dimensional world. One of the difficulties is to visualize four-dimensional space, as we observe a three-dimensional world through our visual system and through information processing in the brain which makes us conscious of it.

In Minkowski's approach [7] to four-dimensional space there exist the speed of *light* as a factor of four-dimensional space. It is possible to write this as

$$x_1^2 + x_2^2 + x_3^2 - c^2 t^2 = 0. \tag{3}$$

This means, having in mind *Fig. 1b* and *Table I*, that *ct* is dimension $N = \bar{1}$. But from our results given in Table I there must be one more relationship between the three-dimensional space (x_1, x_2, x_3) and time (t). From a dimensional point of view we can write a four-dimension as follows:

$$c \,.\, t(=)\, m \,.\, \frac{s}{s}(=)\, \frac{m}{s} \,.\, s \tag{4}$$

and there is only one more possibility for a relation of one of three dimensions (x_1, x_2, x_3) and time (t) as

$$m \,.\, \frac{s}{s}(=)\, m \,.\, s\frac{1}{s}(=)\, \kappa_\kappa \,.\, F \tag{5}$$

where: κ_κ is a *cardinal dimensional measure* as a *main information code* of space — time structures and f — *frequency*.

Now we can write:

$$x_1^2 + x_2^2 + x_3^2 - (ct)^2 - (\kappa_\kappa \,.\, f)^2 = 0 \tag{6a}$$

or

$$x_1^2 + x_2^2 + x_3^2 = (ct)^2 + (\kappa_\kappa \,.\, f)^2 \tag{6b}$$

and we can see that it is equivalent to Fig. 1, where $N = \bar{1}$ is *ct* and $N = \bar{2}$ is $\kappa_\kappa \,.\, f$.

Dimension $N = \bar{2}$ is the product of *cardinal dimensional measure* as a unity of space-time, and *frequency*. According to equation (5) κ_κ has to have the opposite meaning of velocity in the classical sense, because it represents the *space-time code* as a state of *rest* being the same value as *light velocity*, as an *invariant measure* of space-time.

2. Information Physics

According to our results, from both information and physics point of view there exist the realities which can be called *Holopent*, marked $H(5)$ $H(3/2)$. In Greek *holos* means entire, and *pente* means five, so we named this reality *Holopent*, since its base is five-dimensional.

a) We will define *consciousness* as the informational state of Holopent, marked *cH*. It should be noticed that there exist $cH(5)$ and $cH(3/2)$. $cH(5)$ is defined as mapping:
$cH(5):H(5) \rightarrow \kappa(5°)$, and $cH(3/2)$ is defined as mapping:
$cH(3/2)$: $cH(3/2) \rightarrow cH(3/\bar{2})$, while there are two levels:

First level: $$cH(3/\bar{2})_{N=\bar{1}=4} = cH^4 \tag{7a}$$

Second level: $$cH(3/\bar{2})_{N=\bar{2}=5} = cH^5 \tag{7b}$$

where we define:

the *perceptive* consciousness *cpH* as:

$$cp_1H{:}\mathbf{N}(3) \rightarrow cH_4 \text{ — lower level}$$
$$cp_2H{:}\mathbf{N}(3) \rightarrow cH_5 \text{ — higher level} \tag{8}$$

the *non-perceptive* consciousness *cucH* as:

$$cucH{:}\ \mathbf{N}(\bar{1}) \rightarrow cH^5 \tag{9}$$

the *self-consciousness csH* as:

$$csH{:}cH(3/2) \rightarrow cH(3/\bar{2}) \tag{10}$$

b) We will define neurocomputing (*nC*) as:
biological:

$$nb_1C{:}\ \mathbf{N}(3) \rightarrow \mathbf{N}(\bar{1})_{\bar{1}=4} \text{ — external} \tag{11a}$$

$$nb_2C{:}\ \mathbf{N}(4) \rightarrow \mathbf{N}(\bar{1})_{\bar{1}=4} \text{ — internal} \tag{11b}$$

artificial:

$$na_1C{:}\ \mathbf{N}(\bar{1}) \rightarrow \mathbf{N}(2) \text{ — first level} \tag{11c}$$

$$na_2C{:}\ \mathbf{N}(\bar{2}) \rightarrow \mathbf{N}(3) \text{ — second level} \tag{11d}$$

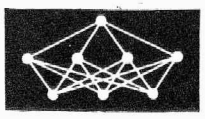
$$na_3C{:}\ \mathbf{N}(4) \rightarrow \mathbf{N}(\bar{3}) \text{ — third level} \tag{11e}$$

From the aspect of *informational physics*, it is possible to realize *artificial consciousness* only as:

$$\text{a) } \mathbf{N}(4) \,.\, \mathbf{N}(\bar{3}) = \mathbf{N}(0) \tag{12}$$

while in the base of biological consciousness is the relation:

$$\mathbf{N}(2) \,.\, \mathbf{N}(\bar{1}) = \mathbf{N}(0) \text{ — first level}$$
$$\mathbf{N}(3) \,.\, \mathbf{N}(\bar{2}) = \mathbf{N}(0) \text{ — second level} \tag{13}$$



because the expressions (12), and (13) give $\mathbf{N}(0) = H(3/2)$ which we defined as the point of departue of consciousness.

Relation with $\mathbf{N}(5)$ have no meaning because the $\mathbf{N}(5)$ is optimal by itself. In the negative dimension $N = \bar{3}, \bar{4}, \ldots$ dimension values 6, 7, 8, ... appear, which means that dimensions greater than N = 5 also have no meaning.

2.1. Information Physics and Biological Neurocomputing

In the scope of biological neurocomputing there exist two main possibilities given in relations (11a) and (11b). The contemporary approach of neurocomputing (pattern recognition, vision, attention, etc.) is given in relation (11a), while the relation (11b) expresses the neurocomputing which leads to the first level of consciousness as:

$$nb_2C\colon \mathbf{N}(4) \rightarrow \mathbf{N}(\bar{1}) \rightarrow cH^4 \tag{14}$$

i. e. real physical processes $\mathbf{N}(4)$ map themselves into the space-time structures. We will define information $I(R^3, t)$ as a space-time pattern in (x_1, x_2, x_3, t)-space which is the true picture of the real physical world into the $\mathbf{N}(4)$. Having in mind that the dimensions in *Table I* are reciprocal {($\mathbf{N}(2) - \mathbf{N}(\bar{1})$, $\mathbf{N}(3) - \mathbf{N}(\bar{2})$, $\mathbf{N}(4) - \mathbf{N}(\bar{3}) \ldots$}, and so are, accordingly, their spaces, we will take the Fourier-space, as a reciprocal space, for the connection between *reality* and the *picture of reality*.

From the aspect of informational physics we can define the Fourier-space by using the variables $\mathbf{k}$ and ω where:

$$|\mathbf{k}| = \frac{2\pi}{\lambda}, \; \omega = 2\pi\nu \tag{15}$$

where λ is wavelength and ν frequency of the wave.

By using Fourier transformation it is possible to write [8]:

$$I(R^3, t) = \int_{-\infty}^{\infty} \mathrm{F}(\mathbf{k}, \omega)\, exp\,\{\mathrm{i}[\mathbf{k} \cdot R^3 - \omega t]\} \frac{\mathrm{dk}}{(2\pi)^3} \cdot d\omega \tag{16}$$

and in that way, in the scope of the $\mathrm{F}(\mathbf{k}, \omega)$ space, it is possible to measure the values which determine the information $I(R^3, t)$, as a space-time pattern.

On the other hand, we can bring together the variables $\mathbf{k}$, ω with the energy (E) and momentum (p) in the following way:

$$E = h\omega, \; p = h\mathbf{k}. \tag{17}$$

On the basis of energy and momentum, we define $\mathbf{N}(4)$ — space as $\mathbf{N}(4) = (p, E)$ — space where p has components p_1, p_2, p_3.

By further application of the quantum theory [8] it is possible to arrive at Table II which presents the neurocomputing as a mapping of reality into its own image and the appearance of the consciqusness of the reality, because it is possible to write the expression (16) in the following way:

$$I(R^3, t) = \psi(r, t) = h^{\frac{1}{4}} \int_{-\infty}^{\infty} \psi(p, E) \exp$$

$$\left\{\mathrm{i}\left[\frac{p}{h} r - \frac{E}{h} t\right]\right\} \frac{\mathrm{d}p}{(2\pi)^3} \mathrm{d}E \tag{18}$$

and the inversive information gives:

$$\mathbf{N}(4) = \psi(p, E) = \int_{-\infty}^{\infty} \psi(r, t)$$

$$\exp\left\{-\mathrm{i}\left[\frac{p}{h} r - \frac{E}{h} t\right]\right\} \mathrm{d}r \frac{\mathrm{d}t}{(2\pi)} \tag{19}$$

The connection which is realized between $I(R^3, t)$ and $\mathbf{N}(4)$ is established upon the following relations:

$$-\mathrm{i}h \frac{\partial}{\partial r} \psi(r, t) = h^{\frac{1}{4}} \int_{-\infty}^{\infty} p\psi(p, E) \exp\left\{\mathrm{i}\left[\frac{p}{h} r - \frac{E}{h} t\right]\right\}$$

$$\frac{\mathrm{d}p}{(2\pi)^3} \mathrm{d}E \tag{20}$$

where:

$$\frac{\partial}{\partial r} = \mathrm{i}\frac{\partial}{\partial x_1} + \mathrm{j}\frac{\partial}{\partial x_2} + \mathrm{k}\frac{\partial}{\partial x_3}, \tag{21}$$

and operator $\hat{\mathrm{p}}$ which is equivalent to the momentum p:

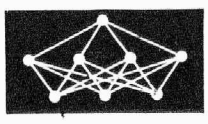
$$\hat{\mathrm{p}} = -ih\frac{\partial}{\partial r} \text{ or } \hat{\mathrm{p}}_{\mathrm{i}} = -\mathrm{i}h\frac{\partial}{\partial x_{\mathrm{i}}}; \; \mathrm{i} = 1, 2, 3. \tag{22}$$

Similarly it is possible to write equation (19) in the following way:

$$-\mathrm{i}h\frac{\partial}{\partial r} \psi(\pi, E) = \int_{-\infty}^{\infty} p\psi(\rho, \tau)$$

$$exp\left\{i\left[\frac{p}{\hbar} r - \frac{E}{h} t\right]\right\} \mathrm{d}r \frac{\mathrm{d}t}{(2\pi)} \tag{23}$$

where the operator $-\mathrm{i}h\frac{\partial}{\partial E}$ is equivalent to time, so that in the brain, as the quantum mechanical machine, an image on the *external world* (x_1, x_2, x_3, t) is formed through the sensor system, but also the picture of the non-perceptive world on the *internal world* $(-\mathrm{i}h \,.\, \partial/\partial x_{1,2,3}$ and $\mathrm{i}h \,.\, \partial/\partial t)$ the *images* of which can be seen during the sleep or in some altered states of conciousness.



2.2 Information Physics and Artificial Neurocomputing

In the scope of the artificial neurocomputing (expression (11e)) which can bring to the artificial consciousness (expression (12)) it is possible to define the space $\mathbf{N}(\bar{3}) = (R^3, E)$ — space and its image $\mathbf{H}(\bar{3}) = (p, t)$ space-time structure as it is given in the *Table III.*

It is necessary to introduce the following state value in order for the fourth dimension to exist through the energy:

$$x_4 = E \cdot \frac{1}{F}(=)\ \{\mathrm{kg}\ \frac{\mathrm{m}}{\mathrm{s}^2} \cdot \mathrm{m}\} \cdot \{\frac{\mathrm{s}^2}{\mathrm{kg} \cdot \mathrm{m}}\}\ (=)\ \mathrm{m} \qquad (24)$$

which is reciprocal to the force, and which can be expressed by $\frac{1}{F} = F\alpha$. In the context of the relationship between space and time, new value in $\mathbf{N}(\bar{3})$ can only appear as the combination c from $\mathbf{N}(\bar{1})$ and f from $\mathbf{N}(\bar{2})$, as a new value: $C_f\ (=)\ c(\chi_0)\ \frac{\mathrm{m}}{\mathrm{s}} \cdot \frac{1}{\mathrm{s}}\ (=)\ \frac{\mathrm{m}}{\mathrm{s}^2}$ which shows that the inversion in the expression (24) is acceleration of gravity. $C_f\ (=)\ c(\chi_0)$ if and only if $f = 1\mathrm{Hz}$, where χ_0 is here cardinal logical operator, and c is space-time invariant. In other words, the coupling of the *consciousness* with the *mass* is possible on the principle of inversion of *gravity*, similarly as the coupling of the *consciousness* with the electricity is achieved in biological systems. The relation between the biological and artificial consciousness on the basis of the informational physics is given in the *fig. 2.*

3. Relativistic Theory of Information

It was shown in (3) that K_κ from the expression (5), as a cardinal information code, is a space-time pattern with the value $3 \cdot 10^{10}$ [cms]. Since this entity (κ_κ) participates in cH^5 (expression 7b) and $cucH$ (expression 9), i. e. in determination of the second level of our consciousness, and in mapping of the contents from the first level of the consciousness into the second level, then it necessarily means that there is a direct connection between the *consciousness* cH^5 and the biophysical state of κ_κ. According to equation (6b) the *frequency* as the co-factor H^5 which gives different states of consciousness. First assumptions on the connection between the consciousness and electro-magnetic waves, based on intuition and without the knowledge of the nature of this connection, are given in the studies [9, 10].

A cardinal information code κ_κ as a property of *Holopent* $H(3/\bar{2})$ according to equations (6a) and (6b) directly correlates with speed of light and biomolecular and brain frequency. In other words basic molecular information "hardware" of the brain and its frequency are coupling.

Table II: Parameters of the biological neurocomputing (Adapt. from Ref. 8).

REALITY $\mathbf{N}(4) = (p, E)$	PICTURE OF REALITY IN THE BRAIN $cH^4 = \mathbf{I}(R^3, t)$
$i\hbar \frac{\partial}{\partial p_1}$	x_1
$i\hbar \frac{\partial}{\partial p_2}$	x_2
$i\hbar \frac{\partial}{\partial p_3}$	x_3
$-i\hbar \frac{\partial}{\partial E}$	t
p_1	$-i\hbar \frac{\partial}{\partial x_1}$
p_2	$-i\hbar \frac{\partial}{\partial x_2}$
p_3	$-i\hbar \frac{\partial}{\partial x_3}$
E	$i\hbar \frac{\partial}{\partial t}$

Table III: Parameters of the artificial neurocomputing which leads to the phenomenon of the artificial consciousness (Adapt. from Ref. 8).

REALITY $\mathbf{N}(\bar{3}) = (R^3, E)$	PICTURE OF REALITY IN ARTIFICIAL BRAIN $\mathbf{I}(p_i, t) = \mathbf{H}(\bar{3})$
x_1	$i\hbar \frac{\partial}{\partial p_1}$
x_2	$i\hbar \frac{\partial}{\partial p_2}$
x_3	$i\hbar \frac{\partial}{\partial p_3}$
$-i\hbar \frac{\partial}{\partial E}$	t
$-i\hbar \frac{\partial}{\partial x_1}$	p_1
$-i\hbar \frac{\partial}{\partial x_2}$	p_2
$-i\hbar \frac{\partial}{\partial x_3}$	p_3
E	$i\hbar \frac{\partial}{\partial t}$

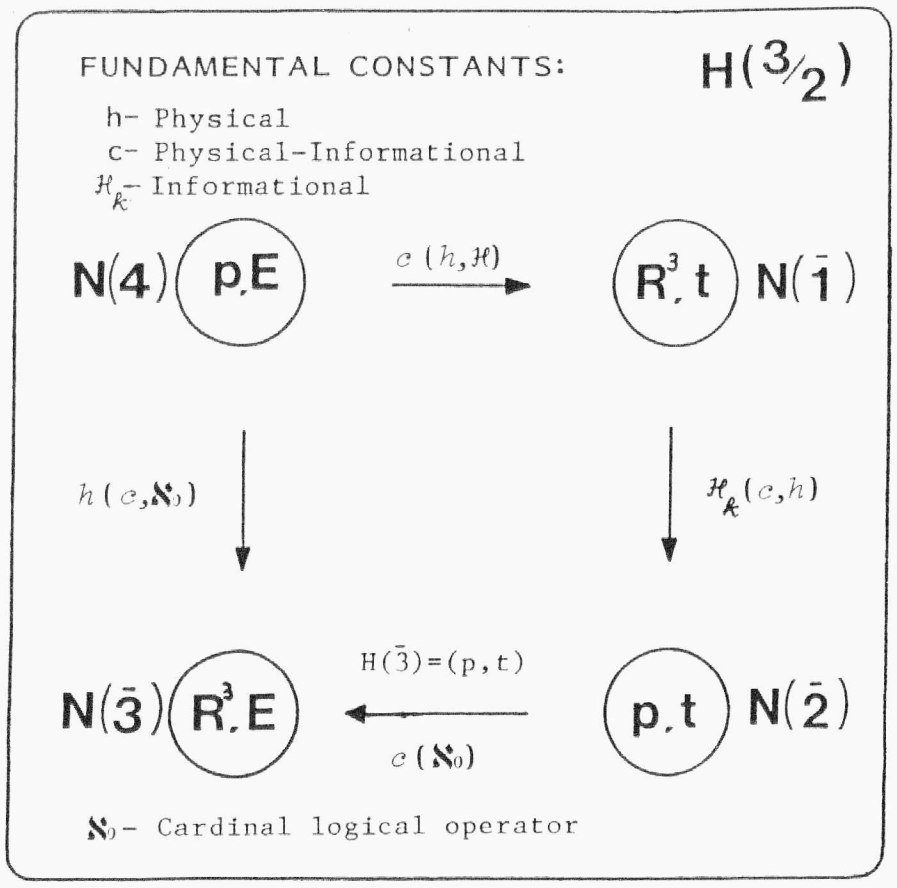


Fig. 2 Connection between the natural and artificial intelligence based on information physics. The Holopent H(3/2) is in the background as the non-manifested mass. The physical reality is the 4-dimensional space **N***(4)-space. The first realization of the* **N***(4) space is the lower level of the biological consciousness in* **N***(1̄), and further realisation into* **N***(2̄), as a higher level of the biological consciousness. Interaction of* **N***(4) and* **N***(2̄) gives* **N***(3̄) which is the artificial consciousness. As the point of departure of this process is in H(3/2) which is based on the laws the golden mean, (see Ref. 3), then the origin of* **N***(1̄),* **N***(2̄) and* **N***(3̄) and their information codes must follow the same rules.*

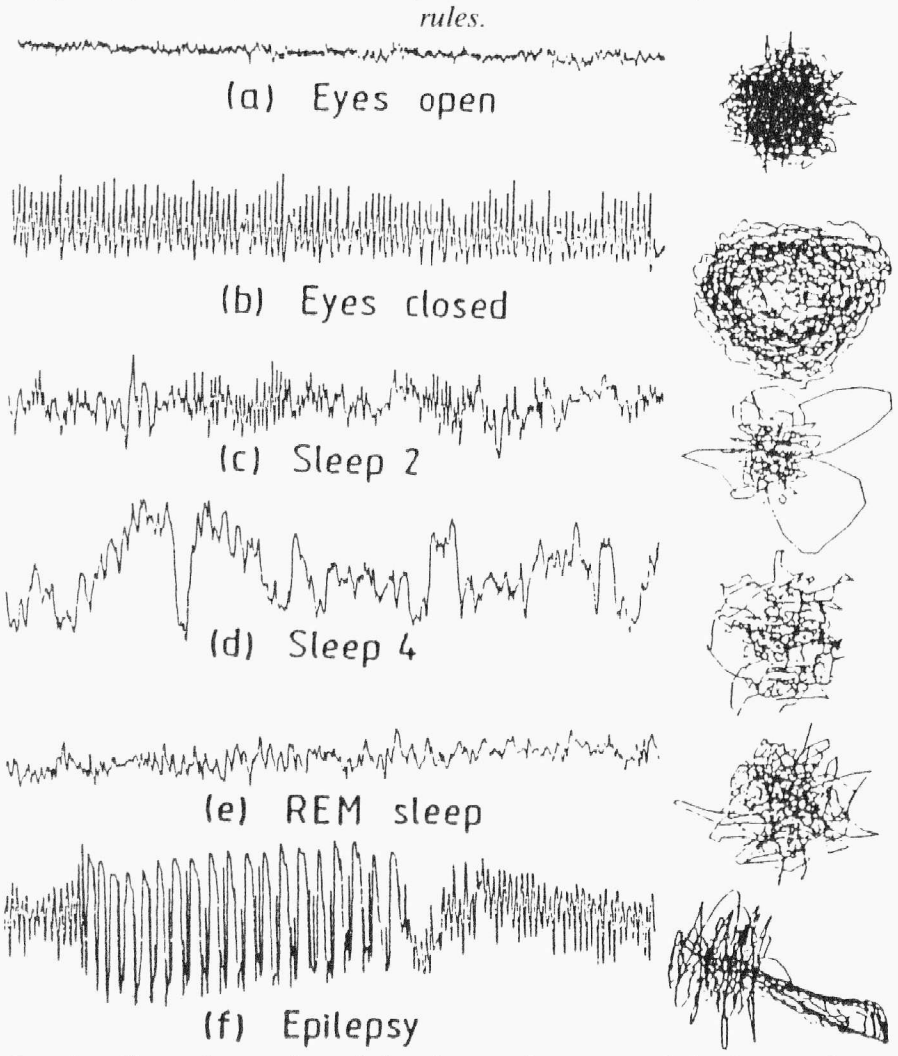


Fig. 3. Typical episodes of the electrical activity of the human brain as recorded from the electroencephalogram (EEG) together with the corresponding phase portraits. These portraits are the two-dimensional projections of three dimensional constructions. The EEG was recorded on a FM analog tape and processed off-line (signal digitized in 12 bits, 250 Hz freq., 4th order 120 Hz low pass filter). [After ref. 11]

From our point of view a very useful quantity for the characterisation of brain dynamic activities as Holopent $H(3/\bar{2})$ is EEG. This is one of the techniques for recording the electrical activity of the brain and if we record EEG together with phase portraits, as a two-dimensional projection of the three-dimensional constructions, it is possible to find a relation between states of EEG and $H(3/\bar{2})$ through fractals [3].

Figure 3 illustrates typical episodes of the electrical activity of the human brain as recorded from the electroencephalogram. *Figure 4* presents the hierarchy of brain states in function of their magnitude.

a)

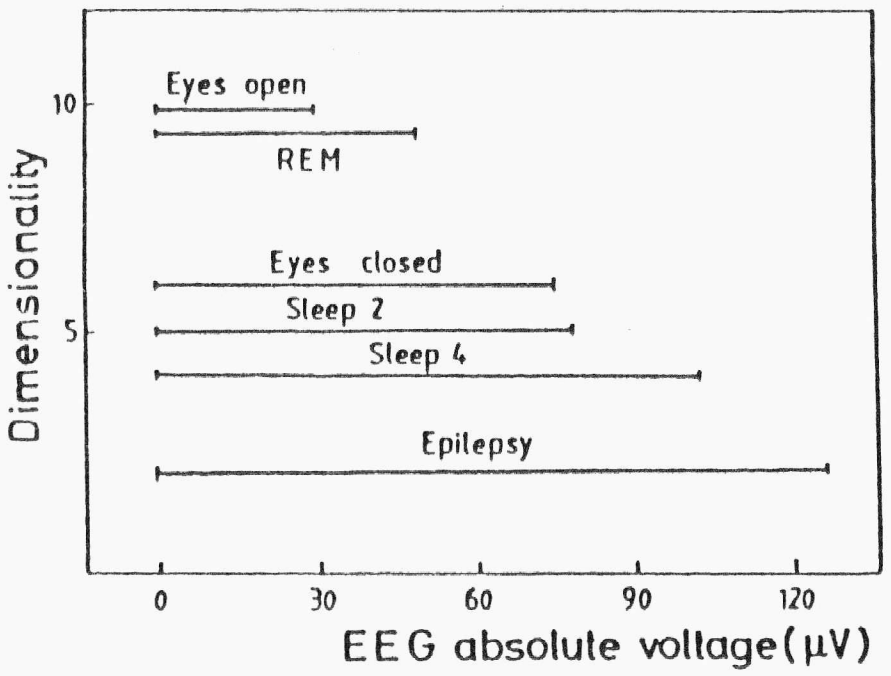


Fig. 4a Representation of the hierarchy of brain states in function of their magnitude. An obvious relation is seen between the EEG variability, quantified by its dimensionality, and the EEG synchrony as reflected by the magnitude of the voltage. [After ref. 11]

b)

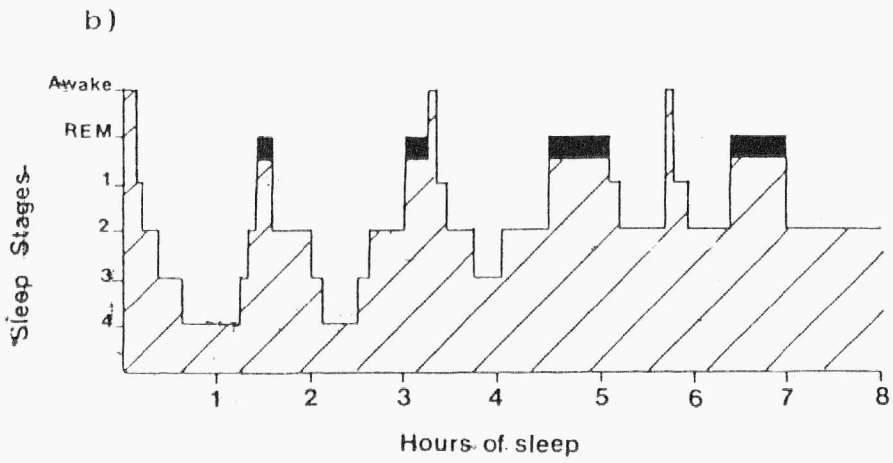


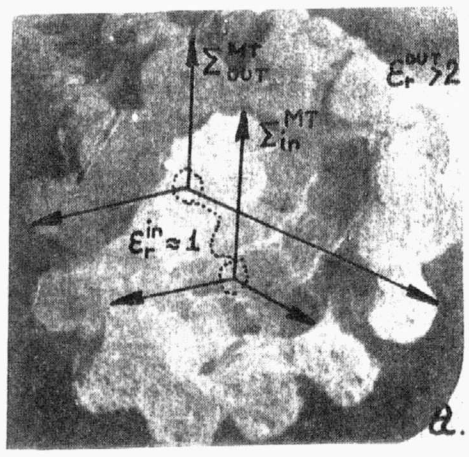
Fig. 4b A simplified "hypnogram" of sleep stage changes over the night in young human adults. [After ref. 12]

A very significant problem, closely related to consciousness, is subjective time sense. From everyday experiences it is known that subjective time sense depends on our psycho-physiological state. So, many people have sense that the objective time in childhood has elapsed slower than in adulthood. Even more striking dilatations of subjective time sense have been observed in altered states of consciousness (REM sleep phase, hypnosis, meditation, the psychedelic drugs influence and some psychopathological states, and near-death experiences).

In the frame-work of the model the "subjective" reference frame will be attached to the electromagnetic component of the scanning brainwaves, and the "objective" to the structure of κ_k (cardinal information code) which is in each neuron of the brain. In fact, the "subjective" reference frame will be attached to those brainwaves whose informational content refer to individual "self". It is understood that the informational content of the individual "self" is simultaneously excited (from the brain's structure with κ_k properties — DNA, MT and neural network, to the brain waves) every time when any new information or sensation is excited.

A physical mechanism that can account for the striking dilatations of available subjective time is the relativistic one, if only consciousness can be associated with κ_k (cardinal information code) as a reference frame in the neuron and brain. Such a "subjective" reference frame could be only associated with an electromagnetic component of brainwaves, which are generated by microtubule waves (as a part of κ_k) and their ionic currents inside the MT, through collective action of a great many of the neurons in the brain.

Microtubules are cell organelles with the outer diameter around 30 nm and inner diameter around 14 nm *(Fig. 5a)*. The outer layer of the microtubule is exposed to the effect of the ions from the cytoplasm which affects the changes of the mass and the dipole moment of the subunits. Due to the change of the dipole moment and the mass of the subunits, MT oscillate with the following electromagnetic and accustic frequencies: $f_{EM} \approx 6 \cdot 10^{15}$ Hz and $f_{AC} \approx 10^{10}$ Hz [13]. The ion currents of very low intensity 10—100 nA may appear inside the microtubule, with the very low concentration of ions which gives relative dialectric permitivity: $\varepsilon_r = 1 + (10^{-10} - 10^{-15})$. Having this in mind we can say that the speed of flow of the ion current [14] in the microtubule is



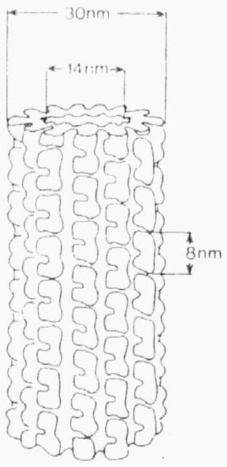


Fig. 5a Diagram of the basic structure of a normal eucaryotic microtubule: as described in the text, most microtubules are assembled from 13 longitudinal protofilaments, each made up of polar tubulin dimers. A dimer consists of two related monomer subunits of about equal molecular weight (55 000) but with specific differences in amino acid sequence. The dimensions shown are those determined for hydrated specimens by X-ray diffraction. [Adapt. from ref. 18]

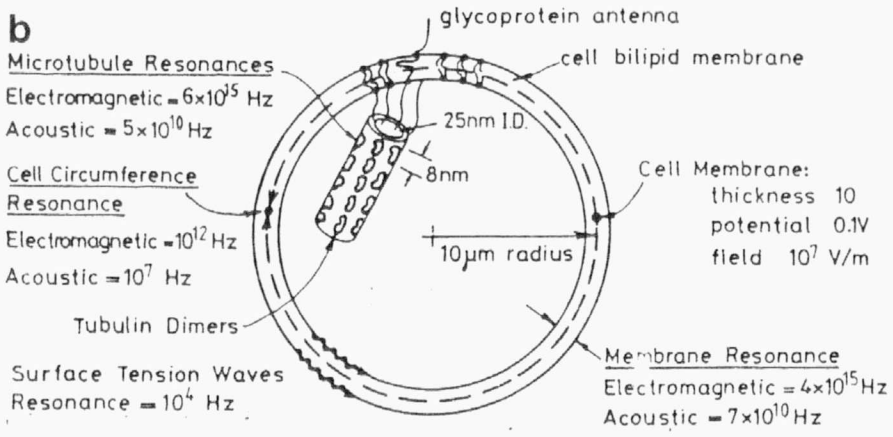


Fig. 5b This figure is a physical model of an "ideal" biological cell showing its possible resonances. Since most biomolecules are electrical dipoles they will behave like microphones turning acoustic waves into electrical waves, and like loud-speakers turning electrical waves into acoustic waves so, the whole cell will act as an oscillating interacting entity [After ref. 13]

$$v = \frac{c}{\sqrt{\varepsilon_r}} = \text{const.} \tag{25}$$

Relativistic relation between the frequencies [15, 16] measured in the two reference frames, moving away from one to another ($\alpha = \pi$), it is possible to write in this form:

$$f_{in}^{MT} = f_{out}^{MT} \cdot \left. \frac{\sqrt{1 - \frac{v^2}{c^2}}}{1 - \frac{v}{c} \cdot \cos\alpha} \right|_{\alpha = \pi} = f_{out}^{MT} \cdot \frac{\sqrt{1 - \frac{1}{\varepsilon_r}}}{1 + \frac{1}{\sqrt{\varepsilon_r}}} \tag{26}$$

According to the equation (26) it is possible to calculate the frequency which is implemented inside the microtubule on the basis of the electromagnetic waves:

$$f_{in}^{MT} = 6 \cdot 10^{15} \cdot \frac{\sqrt{1 - \frac{1}{\varepsilon_r}}}{2} \tag{27}$$

which for values $\varepsilon_r = 1 + (0{,}5 - 20 \cdot 10^{-15})$ gives the frequwency range from 1 — 60 Hz. Having in mind the sub-neural factor MT for the neural metworks [17] and the work of the brain on the whole, it can be said that the electromagnetic waves of the brain (EEG) originate from the oscillatory processes of microtubules and ion currents which form in them on the basis of the relativistic phenomena $f^{Brain} = f_{in}^{MT}(c, \kappa_k, \varepsilon_r)$.

Taking into account all that we said above, we can write:

$$\Delta t^{sub} = \left. \frac{\Delta t_o^{obj}}{\sqrt{1 - \frac{1}{\varepsilon_r}}} \right|_{\varepsilon_r \approx 1} \gg \Delta t_o \tag{28}$$

where Δt_o^{obj} is the real physical time which gives κ_k (cardinal informational code), and Δt^{subj} — time which is implemented inside the MT as a whole acquires this property through the multitude of neurons.

4. Conclusion

In this paper we considered the phenomenon of consciousness and its connection with the neurocomputing from both aspects: fundamental physical laws bases on the quantum field theory, and neurcomputing based on the toplogical-geometric approach. It has been noticed that the consciousness as the global property of the brain has its point of departure on the molecular level. Phenomena that appear on the relation: sub-neural osccillatory processes — brain, are based on the relativistic phenomena. This shows that the informational physics applied to the biological systems is actually relativistic. These results throw new light on the problem of the subjective, and open up new field — relativistic cybernetics, as a science based on informational physics.

As a result of research it is conclusive that *wave-particle* phenomena exist as do pure *wave* and *pure code of* $\kappa(5°)$. As $H(3/\bar{2})$ is in relation to $H(3/2)$ through c and κ_k, and $H(3/2)$ being related to $\kappa(5°)$, then in c (light) there must exsist *pure wave* devoid of energy and momentum, concluding that in structures obtained in κ_k (MT, DNA) there must exist *pure code* devoid of mass. Experimental technique for both entities could be identified, this being one of the objects of our future research.

References

[1] M. Heidegger: An Introduction to Metaphysics, Yale University Press, 1987.

[2] D. Koruga: Biocomputing, HICS-24 IEEE Computer Society Press, 269—275, (1991).

[3] D. Koruga: Neurocomputing: A geometric-topological approach, in book: Ed. M. Novak, E. Pelikan, NEURONET-90, World Scientific Pub., 1991 (in press).

[4] R. Hamming: Coding and Information Theory, Pentice—Hall, 1986.

[5] T. Kohonen: Self-organization and Associate Memory, Springer-Verlag, 1988.

[6] H. L. Ryder: Quantum Field Theory, Cambridge University Press, 1985.

[7] S. W. Hawking, G. F. R. Ellis: The large scale structure of space-time, Cambridge University Press, 1973.

[8] W. Schommers: Space-time and Quantum Phenomena, in book: Ed. W. Schommers, Quantum Theory and Pictures of Reality, Springer-Verlag, 1989.

[9] D. Raković, D. Koruga, Z. Martinović and G. Stanojević: Molecular electronics and neural networks: Significance of ionic structure, Proc. Ann. Int. Conf. IEEE/EMBS, **11**, 1136—1367, 1989.

[10] D. Raković, D. Koruga, D. Djaković, Z. Martinović, V. Desimirović, and D. Minić: Ultralow frequency "optical" biocomputers: Biophysical arguments, in book: Ed. F. Hong, Molecular Electronics: Biosensors and Biocomputers, Plenum Press, 1989.

[11] A. Babloyantz and A. Destexhe: Chaos in Neural Networks, IEEE Conf. on Neural Networks, **IV**, 31—39, 1987, San Diego.

[12] J. Horne: Why We Sleep, Oxford University Press, 1990.

[13] C. W. Smith and S. Best: Electromagnetic Man, J. M. Dent & Sons Ltd. 1989.

[14] F. F. Chen: Introduction to Plasma Physics, Plenum Press, 1974.

[15] L. D. Landau and E. M. Lifschics: Field Theory, Tom II, 1988 (in Russian).

[16] D. Raković: Biophysical Basic of Consciousness and Neural Networks, In Proc. ECPD Neurocomputing **1**, No. 1; 78—91, 1990, Ed. D. Koruga.

[17] D. Koruga: Molecular networks as a sub-neural factor of neural networks, BioSystems, **23**, 297—304, 1990.

[18] Ed. K. Roberts and J. S. Hyans: Microtubules, Academic Press, 1979.

Book Review:

Wasserman P. D.: Neural Computing: Theory and Practice
Van Nostrand Reinhold Co., New York, NY U. S. A, 1989, 230pp., ISBN 0-442-20743-3

This is one of the first book on neural networks theory. It covers main paradigms used today: perceptrons, backpropagation, counterpropagation, Boltzmann machines, Hopfield nets, bidirectional associative memories, adaptive resonance theory (ART), optical and biological networks. Neither the topics nor their presentation are original, the book is (deliberately) repetitious (each chapter is intended to be self-contained), used formalism is rather simple (with little mathematics involved) and some parts are not quite well elaborated, stressing more description than explanation (e. g. when treating backpropagation). It is often difficult to conclude, which model is suited for particular application and which problems you really meet (in this way the „Practice" from —the title has limited meaning). On the other hand it represents a good introduction to specific models with general assessment of their properties and contains many concepts relevant to the topics like stability, local minima problem, parameter setting, classification of various learning strategies etc. It also provides basic references for further study. In general, it is a good introduction for a beginner, assuming that he/she would continue, reading at least some items listed in the bibliography and/or some more advanced monograph [like Hecht-Nielsen's — see p. 57] to get more information on network composition, higher mathematics considerations and real applications.

J. Hořejš

Books Alert

The following books can be interesting for the readers of our Journal:

Analog VLSI Implementation of Neural Systems. Ed. Carver Mead and Mohammed Ismail. — Boston, MA: Kluwer Academic, 1989, 248 pp., bound, $ 55. 00, ISBN 0-7 9923-90407.

The contents are as follows: "A Neural Processor for Maze Solving"; „Resistive Fuses: Analog Hardware for Detecting Discontinuities in Early Vision"; „CMOS Integration of Herault-Jutten Cells for Separation of Sources"; „Circuit Models of Sensor Transduction in the Cochlea"; "Issues in Analog VLSI and MOS Techniques for Neural Computing"; "Design and Fabrication of VLSI Components for a General Purpose Analog Neural Computer"; "A Chip that Focuses an Image on Itself"; "A Foveated Retina-Like Sensor Using CCD Technology"; "Cooperative Stereo Matching Using Static And Dynamic Image Features"; "Adaptive Retina."

LEARNING IN A PARTIALLY HARD-WIRED RECURRENT NETWORK

C.-M. Kuan), K. Hornik**)*

Abstract:

In this paper we propose a partially hard-wired Elman network. A distinct feature of our approach is that only minor modifications of existing on-line and off-line learning algorithms are necessary in order to implement the proposed network. This allows researchers to adapt easily to trainable recurrent networks. Given this network architecture, we show that in a general dynamic environment the standard back-propagation estimates for the learnable connection weights can converge to a mean square error minimizer with probability one and are asymptotically normally distributed.

1. Introduction

Neural network models have been successfully applied in a wide variety of disciplines. Typically, applications of networks with at least partially modifiable interconnection strengths are based on the so-called multilayer *feedforward* architecture, in which all signals are transmitted in one direction without feedbacks. In a dynamic context, however, a feedforward network may have difficulties in representing certain sequential behavior when its inputs are not sufficient to characterize temporal features of target sequences (Jordon, 1985). From the cognitive point of view, a feedforward network can perform only passive cognition, in that its outputs cannot be adjusted by an internal mechanism when static inputs are present (Norrod, O'Neill, & Gat, 1987). These deficiencies thus restrict the applicability of feedforward neural network models in dynamic environments.

In view of these problems, researchers have recently been studying *recurrent* networks, i.e., networks with feedback connections, see e.g., Jordon (1986), Elman (1988), Williams & Zipser (1988), and Kuan (1989). In a recurrent network, recurrent variables compactly summarize the past information and, together with other input variables, jointly determine the network outputs. Because recurrent variables are generated by the network, they are functions of the network connection weights. Owing to this parameter dependence, the standard back-propagation (BP) algorithm for feedforward networks cannot be applied because it fails to take the correct gradient search direction (cf. Rumelhart, Hinton & Williams, 1986). Kuan, Hornik & White (1990) propose a recurrent BP algorithm generalizing the standard BP algorithm to various recurrent networks. However, this algorithm has quite complex updating equations and restrictions, and therefore cannot be used straightforwardly by recurrent networks practitioners.

In this paper we suggest an easier way to implement recurrent networks. We focus on a variant of the Elman (1988) network, in which only a subset of hidden unit activations serve as recurrent variables. We propose to hard-wire the connections between the recurrent units and their inputs. This approach has the following advantages. First, the resulting network avoids the aforementioned problem of parameter dependence. Second, the necessary constraints on recurrent connections suggested by Kuan, Hornik, & White (1990) can easily be imposed by hard-wiring. Third, off-line learning is made possible for the proposed network. Consequently, only minor modifications of existing on-line and off-line learning algorithms are needed. This is very convenient for neural network practitioners. Given this hard-wired network, we show that in general dynamic environments the resulting BP estimates converge to a mean squared error minimizer with probability one and are asymptotically normally distributed. Our convergence results extend the results of Kuan, Hornik, & White (1990) for general recurrent networks and are analogous to the results of Kuan & White (1990) for feedforward networks.

This paper proceeds as follows. In section 2 we briefly review recurrent networks. In section 3 we discuss a variant of the Elman network and its learning algorithms. We establish strong consistency and asymptotic normality of the learning estimates in section 4. Section 5 concludes the paper. Proofs are deferred to the appendix.

*) Prof. Chung-Ming Kuan
Department of Economics
Box 111
330 Commerce Building (West)
1206 South Sixth Street
Champaign, IL 61820
U.S.A.
ckuan@ux1.cso.uiuc.edu

**) Prof. Kurt Hornik
Institut für Statistik und Wahrscheinlichkeitstheorie
Technische Universität Wien
Wiedner Hauptstraße 8—10/1071
A-1040 Wien
Austria
hornik@eiaida.tuwien.ac.at

2. Recurrent Networks

A three layer recurrent network with k input units, l hidden units with common activation function ψ, and m output units with common activation function φ can be written in the following generic form:

$$o_t = \varphi(Wa_t + v)$$
$$a_t = \psi(Cx_t + Dr_t + b)$$
$$r_t = G(x_{t-1}, r_{t-1}, \theta),$$

where the subscript t indexes time, x is the $k \times 1$ vector of network inputs, a is the $l \times 1$ vector of hidden unit activations, o is the $m \times 1$ vector of network outputs, φ and ψ compactly denote the unitwise activation rules in the output respectively hidden layer, and r_t is the $n \times 1$ vector of recurrent variables which is computed through some general function G from the previous input x_{t-1}, the previous recurrent variable r_{t-1}, and

$$\theta = [\text{vec}(C)', \text{vec}(D)', \text{vec}(W)', b', v']',$$

the vector of all network connection weights. (In what follows, $'$ denotes transpose, the vec operator stacks the columns of a matrix one underneath the other, and $|v|$ is the euclidean length a vector v.)

More compactly, the above network can be written as

$$o_t = \varphi(W\psi(Cx_t + Dr_t + b) + v) \quad (1)$$
$$r_t = G(x_{t-1}, r_{t-1}, \theta). \quad (2)$$

That is, the network output is jointly determined by the external inputs x and the recurrent variables r. Clearly, different choices of G yield different recurrent networks. When $r_t = o_{t-1}$ (output feedback),

$$r_t = G(x_{t-1}, r_{t-1}, \theta) = \varphi(W\psi(Cx_{t-1} + Dr_{t-1} + b) + v)$$

and we obtain the Jordon (1986) network. When $r_t = a_{t-1}$ (hidden unit activation feedbacks),

$$r_t = G(x_{t-1}, r_{t-1}, \theta) = \psi(Cx_{t-1} + Dr_{t-1} + b),$$

and we have the Elman (1988) network.

By recursive substitution, (2) becomes

$$r_t = G(x_{t-1}, r_{t-1}, \theta) = G(x_{t-1}, G(x_{t-2}, r_{t-2}, \theta), \theta) =$$
$$= \ldots =: \ell_t(x^{t-1}, \theta),$$

where $x^{t-1} = (x_{t-1}, x_{t-2}, \ldots, x_0)$ is the collection of past inputs. Hence, r_t is a complex nonlinear function of θ and the entire past of x_t. In contrast with external input x_t, we may interpret r_t as "internal" input, in the sense that it is generated by the network. Given a recurrent network, the standard BP algorithm for feedforward networks does not perform correct gradient search over the parameter space because it fails to take the dependence of r_t on the learnable network weights into account. Consequently, meaningful convergence cannot be guaranteed (Kuan, 1989).

Kuan, Hornik, & White (1990) propose a recurrent BP algorithm which, by carefully calculating the correct gradients and including additional derivative updating equations, maintains the desired gradient search property. To ensure proper convergence behavior, their results also suggest some restrictions on the network connection weights. That is, parameters estimates are projected into some "stability" region whenever they violate the imposed constraints. Thus, much more effort is needed in programming appropriate learning algorithms for recurrent networks. Moreover, some of their conditions to ensure convergence of the recurrent BP algorithm are rather stringent.

3. Partially Hard-Wired Elman Network

In this section we suggest an easier way to implement a variant of the Elman (1988) network. As we have discussed in section 2, improper convergence of the learning algorithms is mainly due to the dependence of the internal inputs r_t on the modifiable network parameters. To circumvent this problem, we propose to modify the Elman network as is depicted in *figure 1*.

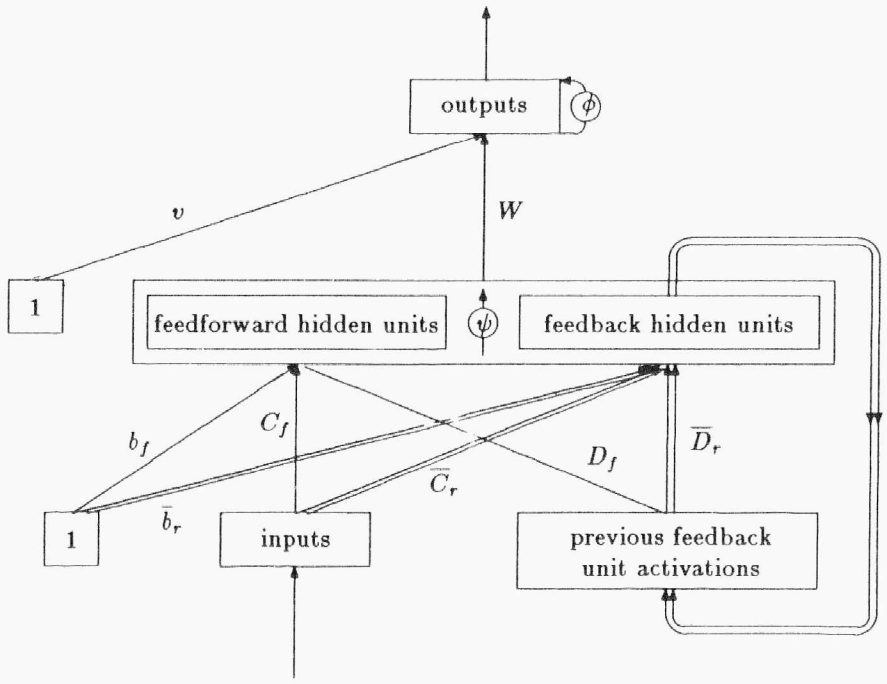


Figure 1. The proposed partially hard-vired recurrent network. Modifiable and hard-wired connections are represented by → respectively ⇒.

The hidden units are partitioned into two groups containing l_f respectively $l_r = l - l_f$ units, and only the units in the second group serve as recurrent units. Intuitively, the units in the first group play the standard role in artificial neural networks, whereas the task of the recurrent units is to "index" information on previous inputs. Furthermore, the connections between the recurrent units and their inputs are hard-wired.

Hence, a is partitioned as $a = [a_f', a_r']'$, where a_f is the $l_f \times 1$ vector of activations of the (purely feedforward) hidden units in the first group, and a_r is the $l_r \times 1$ vector of activations of the feedback (recurrent) hidden units in the second group such that

$$r_t = a_{r,t-1} .$$

If the connection matrices C and D and the bias vector b are partitioned comformably as

$$C = \begin{bmatrix} C_f \\ \overline{C}_r \end{bmatrix}, \quad D = \begin{bmatrix} D_f \\ \overline{D}_r \end{bmatrix}, \quad b = \begin{bmatrix} b_f \\ \overline{b}_r \end{bmatrix},$$

then

$$a_{f,t} = \psi(C_f x_t + D_f a_{r,t-1} + b_f)$$
$$a_{r,t} = \psi(\overline{C}_r x_t + \overline{D}_r a_{r,t-1} + \overline{b}_r) ,$$

where now $\overline{C}_r$, $\overline{D}_r$ and $\overline{b}_r$ are fixed due to hard-wiring. Different choices of $\overline{C}_r$, $\overline{D}_r$ and $\overline{b}_r$ determine how the past information should be represented, hence they are problem-dependent and should be left to researchers.

Hence, writing the proposed network in a nonlinear functional form, we have

$$o_t = \varphi(W\psi(Cx_t + Da_{r,t-1} + b) + v) = \\ =: F(x_t, a_{r,t-1}, \theta) \qquad (3)$$

and

$$r_t = a_{r,t-1} = \psi(\overline{C}_r x_{t-1} + \overline{D}_r a_{r,t-2} + \overline{b}_r) = \\ =: G(x_{t-1}, a_{r,t-2}, \overline{\theta}) , \qquad (4)$$

where now

$$\theta = \left[\mathrm{vec}(\mathrm{W})', \mathrm{vec}(\mathrm{C_f})', \mathrm{vec}(\mathrm{D_f})', \mathrm{b_f'}, \mathrm{v}'\right]'$$

is the $p \times 1$ vector which contains all the learnable network weights, where $p := m(l+1) + l_f(k + l_r + 1)$, and

$$\overline{\theta} = \left[\mathrm{vec}(\overline{\mathrm{C}}_\mathrm{r})', \mathrm{vec}(\overline{\mathrm{D}}_\mathrm{r})', \overline{\mathrm{b}}_\mathrm{r}'\right]'$$

contains all the hard-wired weights. By recursive substitution, (4) becomes

$$r_t = G(x_{t-1}, r_{t-1}, \overline{\theta}) = G(x_{t-1}, G(x_{t-2}, r_{t-2}, \overline{\theta}), \overline{\theta}) = \\ = \ldots = : \mu_t(x^{t-1}, \overline{\theta}) ,$$

cf. equation (2). Thus, $r_t = a_{r,t-1}$ is a function of the entire past of x_t and the hard-wired weights $\overline{\theta}$.

Because r_t is not a function of the learnable weights θ, the aforementioned problem of parameter dependence is thus avoided. It follows that the standard BP algorithm for feedforward networks is applicable to the proposed network with respect to the learnable weigths θ. Letting y_t denote the target pattern presented at time t, the BP algorithm is

$$\hat{\theta}_{t+1} = \hat{\theta}_t + \eta_t \nabla_\theta F(x_t, r_t, \hat{\theta}_t)\,(y_t - F(x_t, r_t, \hat{\theta}_t)) , \quad (5)$$

where η_t is learning rate employed at time t and $\nabla_\theta F$ is the matrix of partial derivatives of F with respect to the components of θ. However, in both theory and practice it is necessary to keep the BP estimates in some compact subset Θ of IR^p, thus preventing the entries from becoming extremely large. This, being a typical requirement in the convergence analysis of the BP type of algorithms, see e.g., Kuan & White (1990) und Kuan, Hornik & White (1990), can, if not automatically guaranteed by the algorithm, be accomplished by applying a projection operator π which maps IR^p onto Θ to the BP estimates. Usually, a truncation device is convenient for this purpose. This requirement entails little loss because it is usually inactive when very large trunaction bounds are imposed.

In light of (5), we only have to modify the existing BP algorithm slightly to incorporate the internal inputs a_r into the algorithm. Furthermore, if a fixed training data set is given, the internal inputs $a_{r,t}$ can be calculated first, and off-line learning methods such as nonlinear least squares can then be applied to estimate the learnable weights θ. These advantages allow researchers to adapt to recurrent networks quite easily. It is then interesting to know the properties of the algorithm (5) applied to the proposed network given by (3) and (4). This is the topic to which we now turn.

4. Asymptotic Properties of the BP Algorithm

Let $\{V_t\}$ be some sequence of random variables defined on a probability space (Ω, F, P), F_r^t be the σ-algebra generated by $V_r, V_{r+1}, \ldots, V_t$, and let $\{Z_t\}$ be a sequence of square integrable random variables on that probability space. We write $E_{t-m}^{t+m}(Z_t)$ for the conditional expectation $E(Z_t \mid \mathrm{F}_{t-m}^{t+m})$ and $\|\cdot\|$ for the norm in $L_2(P)$, i.e., $\|Z\| = (E|Z|^2)^{1/2}$.

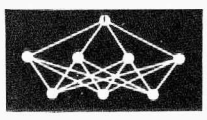
Definition 4.1. Let

$$v_m := \sup_t \| Z_t - E_{t-m}^{t+m}(Z_t) \| .$$

Then $\{Z_t\}$ is *near epoch dependent* (NED) *on* $\{V_t\}$ *of size* $-a$ if for some $\lambda < -a$, $v_m = O(m^\lambda)$ as $m \to \infty$. This definiton conveys the idea that a random variable depends essentially on the information generated by "more or less current" V_t and does not depend too much on the information contained in the distant future or past. The larger the magnitude of the size of v_m, the faster the dependence of the remote information dies out. More details on near epoch dependence can be found in Billingsley (1968), McLeish (1975), and Gallant & White (1988).

The lemma below ensures that recurrent variables are well behaved and do not have too long memory.

Lemma 4.2. *Let* $\{r_t\}$ *be generated by* (4), *where* $\{x_t\}$ *is NED on* $\{n_t\}$ *of size* $-a$ *and the common hidden unit activation function* ψ *is bounded and continuously differentiable with bounded first derivative. If* $|\mathrm{vec}(\overline{\mathrm{D}}_\mathrm{r})| < M_\psi^{-1}$, *where* $M_\psi := \sup_{\sigma \in IR} |\psi'(\sigma)|$, *then* $\{r_t\}$ *is a bounded sequence NED on* $\{V_t\}$ *of size* $-a$.



Remark 1. Notice that if the input data $\{x_t\}$ form a sequence of independent random variables (which is a special case of an NED sequence), then $\{r_t\}$ need not necessarily be mixing but is NED on $\{x_t\}$ of arbitrarily large size, see Gallant & White (1988, pp. 27—31). Hence, introducing the concept of near epoch dependence is not a technical triviality, but a necessity when dealing with feedback networks in stochastic input environments.

In what follows we compactly write the algorithm (5) as

$$\hat{\theta}_{t+1} = \hat{\theta}_t + \eta_t\, h_t(\hat{\theta}_t)\, z_t$$

$h_t(\theta) = \nabla_\theta F(x_t, r_t, \theta)\,(y_t - F(x_t, r_t, \theta))$. Our consistency result is based on the ordinary differential equation (ODE) method of Kushner & Clark (1978), cf. Ljung (1977). This approach is now well-known in analyzing neural network learning algorithms, cf. e.g., Oja (1982), Oja & Karhunen (1985), Sanger (1989), Kuan & White (1990), Kuan, Hornik, & White (1990), and Hornik & Kuan (1990).

We need the following notation. Let $r_0 = 0$ and, for $t \geq 1$, let $r_t := \Sigma_{i=0}^{t-1} \eta_i$. The piecewise linear interpolation of $\{\hat{\theta}_t\}$ with interpolation intervals $\{\eta_t\}$ is

$$\theta^0(r) = \left(\frac{r_{t+1} - r}{\eta_t}\right) \hat{\theta}_t + \left(\frac{r - r_t}{\eta_t}\right) \hat{\theta}_{t+1}, \quad r \in [r_t, r_{t+1}),$$

and for each t, its "left shift" is

$$\theta^t(r) = \theta^0(r_t + r).$$

Observe in particular that $\theta^t(0) = \theta^0(r_t) = \hat{\theta}_t$.

We impose the folowing conditions.

A.1. $\{V_t\}$ and $\{z_t\}$ are defined on a complete probability space (Ω, F, P) such that for some $r \geq 4$,
(i) $\{V_t\}$ is a mixing sequence with mixing coefficients φ_m of size $-r/2(r-1)$ or α_m of size $-r/(r-2)$ and
(ii) the sequence $\{z_t\}$ is NED on $\{V_t\}$ of size -1 with $\sup_t |x_t| \leq M_x < \infty$ and $\sup_t E(|y_t|^r) < \infty$.

A.2. For the network architecture as specified in (3) and (4),
(i) φ and ψ are continuously differentiable of order 3. ψ is bounded and has bounded first order derivative.
(ii) $|\mathrm{vec}(\bar{D}_r)| < M_\psi^{-1}$, where $M_\psi = \sup_{\sigma \in IR} |\psi'(\sigma)|$.

A.3. $\{\eta_t\}$ is a sequence of positive real numbers such that $\sum_t \eta_t = \infty$ and $\sum_t \eta_t^2 < \infty$.

A.4. For each $\theta \in \Theta$, $\bar{h}(\theta) = \lim_t E(h_t(\theta))$ exist.

A.1. allows the data to exhibit a considerable amount of dependence in the sence that they are functions of the (possibly infinite) history of an underlying mixing sequence. For more details on α- and φ- mixing sequences we refer to White (1984). Assuming that the external inputs x_t are uniformly bounded simplifies some technicalities needed to establish convergence and causes no loss of generality, as pointed out by Kuan & White (1990). Desired generality is assured by allowing the y_t sequence to be unbounded. Note that typical choices for ψ such as the logistic squasher and hyperbolic tangent squasher satisfy A.2(i). Condition A.2(ii) is needed in lemma 4.2 and is the constraint suggested by Kuan, Hornik & White (1990) for general recurrent networks. A.3 is a typical restriction on the learning rates for BP types of algorithms. For example, learning rates of order $1/t$ satisfy this condition. A.4 is needed to define the associated ODE whose solution trajectory is the limiting path of the interpolated processes $\{\theta^t(\cdot)\}$.

The result below follows from corollary 3.5 of Kuan & White (1990).

Theorem 4.3. *For the network given by* (3) *and* (4) *and the algorithm* (5), *suppose that assumptions* A.1-A.4 *hold. Then*
(a) $\{\theta^t(\cdot)\}$ *is bounded and equicontinuous on bounded intervals with probability one, and all limits of convergent subsequences satisfy the ODE* $\dot{\theta} = \bar{h}(\theta)$.
(b) *let* Θ^* *be the set of all (locally) asymptotically stable equilibria of this ODE contained in* Θ, *and let* $\mathfrak{D}(\Theta^*) \subset IR^p$ *be the domain of attraction of* Θ^*. *Then, if* $\hat{\theta}_t$ *enters a compact subset of* $\mathfrak{D}(\Theta^*)$ *infinitely often with probability one, and thus in particular, if* $\Theta \subseteq \mathfrak{D}(\Theta^*)$, *then with probability one,* $\theta_t \to \Theta^*$ *as* $t \to \infty$.

Remark 2. Because the elements θ^* of Θ^* solve the equation $\lim_t E(h(z_t, r_t, \theta)) = \bar{h}(\theta) = 0$, they (locally) minimize

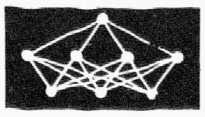
$$\lim E|y_t - F(x_t, r_t, \theta)|^2. \tag{6}$$

Theorem 4.3 thus shows that the BP estimates can converge to a mean squared error minimizer with probability one. Note however that this convergence occurs conditional on $\bar{\theta}$.

Remark 3. By the Toeplitz lemma,
$\lim_T T^{-1} \sum_{t=1}^{T} E|y_t - F(x_t, r_t, \theta)|^2$ is the same as (6). Therefore, the (on-line) BP estimates converge to the same limit as the (off-line) nonlinear least squares estimator.

Remark 4. As y_t is not required to be bounded, our strong consistency result holds under less stringent conditions than those of Kuan, Hornik & White (1990) for the fully recurrent BP algorithm.

To establish asymptotic normality we consider the algorithm (5) with the specific choice $\eta_t = (t+1)^{-1}$. (Note that no limiting distribution results for BP estimators in recurrent networks have been published thus far; in particular, Kuan, Hornik, & White (1990)



give only a consistency result for their recurrent BP algorithm.) Let $U_t := (t+1)^{1/2}(\hat{\theta}_t - \theta^*)$ be the sequence of normalized estimates. The piecewise constant interpolation of U_t on $[0, \infty)$ with interpolation intervals $\{(t+1)^{-1}\}$ is defined as

$$\overline{U}(r) = U_t, \quad r \in [r_t, r_{t+1}),$$

and again, for each t its "left shift" is defined as

$$U^t(r) = \overline{U}(r_t + r), \quad r \geq 0.$$

Finally, let

$$\overline{H}(\theta) := \lim_t E[\nabla_\theta h_t(\theta)] + I_p/2,$$

where I_p is the p-dimensional identity matrix.

Our result follows from the stochastic differential equation (SDE) approach of Kushner & Huang (1979). In contrast with the ODE approach, the interpolated processes can now be shown to converge *weakly* to the solution paths of a corresponding SDE with respect to the Skorohod topology. For more details on weak convergence we refer to Billingsley (1968). The following conditions suffice for the asymptotic normality result.

B.1. A.1(i) holds, and $\{z_t\}$ is a stationary sequence NED on $\{V_t\}$ of size -8 with $\sup_t |x_t| \leq M_x < \infty$ $\sup_t E(|y_t|^8) < \infty$.

B.2. A.2 holds with φ and ψ continously differentiable of order 4.

B.3. $\theta^* \in \text{int}(\Theta)$ is such that $\overline{h}(\theta^*) = 0$ and all eigenvalues of $\overline{H}(\theta^*)$ have negative real parts.

The result below follows from corollary 3.6 of Kuan & White (1990).

Theorem 4.4. *Consider the network given by* (3) *and* (4) *and the algorithm* (5) *with* $\eta_t = (t+1)^{-1}$, *suppose that assumptions* B.1-B.3 *hold and that with probability one,* $\theta_t \to \theta^*$ *as* $t \to \infty$. *Then* $\{U^t(\cdot)\}$ *converges weakly to the stationary solution of the stochastic differential equation*

$$dU(r) = \overline{H}(\theta^* U(r) dr + \overline{\sum}(\theta^*)^{1/2} dW(r),$$

where W denotes the standard p-variate Wiener process and

$$\overline{\sum}(\theta^*) := \lim_t E[h_t(\theta^*) h_{t-j}(\theta^*)'].$$

In particular,

$$(t+1)^{1/2}(\hat{\theta}_t - \theta^*) \xrightarrow{D} N(0, S(\theta^*)),$$

where "$\xrightarrow{D}$" signifies convergence in distribution and

$$S(\theta^*) := \int_0^\infty \exp(\overline{H}(\theta^*)s) \overline{\sum} \exp(\overline{H}(\theta^*)s) ds$$

is the unique solution to the matrix equation $\overline{H}(\theta^*)S + S\overline{H}(\theta^*)' = -\overline{\sum}(\theta^*)$.

Remark 5. If $\eta_t = (t+1)^{-1} R$ is a nonsingular $p \times p$ matrix, the SDE in theorem 4.4 becomes $dU(r) = \overline{H}(\theta^*) U(r) dr + R\overline{\sum}(\theta^*)^{1/2} dW(r)$, and the covariance matrix of the asymptotite distribution of $\hat{\theta}_t$ becomes $RS(\theta^*)R'$.

Remark 6. If the probability that $\hat{\theta}_t$ converges to θ^* is positive, but less that one, the above theorem provides the limiting distribution *conditional* on convergence to θ^*. Hence, if Θ^* contains only finitely many points, assumption B.3 is satisfied for each $\theta^* \in \Theta^*$, and $\hat{\theta}_t$ converges with probability one to one of the elements of Θ^*, then the asymptotic distribution of $\hat{\theta}_t$ is mixture of $N(\theta^*, S(\theta^*))$ distributions, weighted relative to the convergence probabilities.

5. Conclusions

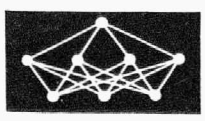
In this paper we propose a partially hard-wired Elman network, in which only a subset of hidden-unit activations is allowed to feed back into the network and connections between these hidden units and input layer are hard-wired. A distinct feature of our approach is that existing on-line and off-line learning algorithms can be slightly modified to implement the proposed network. (Note that off-line learning is not possible for a fully learnable recurrent network.) This is particularly convenient for researchers. Our results also show that the estimates from the standard BP algorithm adapted to this network can converge to a mean squared error minimizer with probability one and are asymptotically normally distributed. These asymptotic properties are analogous to those of the standard and recurrent BP algorithms.

As the convergence results in this paper are conditional on the hard-wired connection weights $\overline{\theta}$, the resulting weight estimates are not fully optimal, in contrast with fully learnable recurrent networks. To improve the performance of the proposed network, one can train the network with various hard-wired connection weights and search for the best performing architecture.

Appendix

Lemma A. *Let* $\{x_t\}$ *be NED on* $\{V_t\}$ *of size* $-a$ *and let the square integrable sequence* $\{r_t\}$ *be generated by the recursion*

$$r_t = G(x_{t-1}, r_{t-1}, \overline{\theta}).$$



Suppose that $G(\cdot, r, \bar{\theta})$ satisfies a Lipschitz condition with $\alpha = 1$ uniformly in r, i.e., there exists a finite constant L such that for all r,

$$|g(x_1, r, \bar{\theta}) - g(x_2, r, \bar{\theta})| \leq L|x_1 - x_2|,$$

and that $G(x, \cdot\, \bar{\theta})$ is a contraction mapping uniformly in x, i.e., there exists some $\rho < 1$ such that for all x,

$$|G(x, r_1, \bar{\theta}) - g(x, r_2, \bar{\theta})| \leq \rho|r_1 - r_2|.$$

Then $\{r_t\}$ is NED on $\{V_t\}$ of size $-a$.

Proof. We first observe that

$$\begin{aligned}
&\|r_t - E_{t-m}^{t+m}(r_t)\| \\
&= \|G(x_{t-1}, r_{t-1}, \bar{\theta}) - E_{t-m}^{t+m}(G(x_{t-1}, r_{t-1}, \bar{\theta}))\| \\
&\leq \|G(x_{t-1}, r_{t-1}, \bar{\theta}) - G(E_{t-m}^{t+m-2}(x_{t-1}), \\
&\quad E_{t-m}^{t+m-2}(r_{t-1}), \bar{\theta})\| \\
&\leq \|G(x_{t-1}, r_{t-1}, \bar{\theta}) - G(E_{t-m}^{t+m-2}(x_{t-1}), r_{t-1}, \bar{\theta})\| + \\
&\quad + \|G(E_{t-m}^{t+m-2}(x_{t-1}), r_{t-1}, \bar{\theta}) - \\
&- G(E_{t-m}^{t+m-2}(x_{t-1}), E_{t-m}^{t+m-2}(r_{t-1}), \bar{\theta})\| \\
&\leq L\|x_{t-1} - E_{t-m}^{t+m-2}(x_{t-1})\| + \rho\|r_{t-1} - \\
&- E_{t-m}^{t+m-2}(r_{t-1})\|,
\end{aligned}$$

where the first inequality follows from the fact that $E_{t-m}^{t+m}(G(x_{t-1}, r_{t-1}, \bar{\theta}))$ is the best mean square predictor of $G(x_{t-1}, r_{t-1}, \bar{\theta})$ among all F_{t-m}^{t+m}-measurable functions and the second inequality follows from the triangle inequality. Hence, we obtain

$$v_{r,m} \leq Lv_{x,m-1} + \rho v_{r,m-1}, \qquad \text{(a1)}$$

where $v_{x,m}$ and $v_{r,m}$ are the NED coefficients for $\{x_t\}$ and $\{r_t\}$, respectively. We must show that for some $\lambda < -a$, $v_{r,m}$ is $O(m^{\lambda})$ as $m \to \infty$. Because $\{x_t\}$ is NED on $\{V_t\}$ of size a, we can find a finite constant C_0 and some $\lambda_0 < -a$ such that $v_{x,m} \leq C_0 m^{\lambda_0}$. By the fact that $\rho < 1$, we can find m_0 and some $\sigma > 1$ such that $\rho\sigma < 1$ and for all $m \geq m_0$,

$$(m/(m+1))^{\lambda_0} \leq \sigma.$$

Let

$$D_0 := \max\left\{\frac{v_{r,m_0}}{m_0^{\lambda_0}}, \frac{C_0 L\sigma}{1 - \rho\sigma}\right\}.$$

We now prove by induction that for all $m \geq m_0$, $v_{r,m} \leq D_0 m^{\lambda_0}$. For $m = m_0$, this is trivially true by the definition of D_0. Suppose we have already shown that for some $m \geq m_0$, $v_{r,m} \leq D_0 m^{\lambda_0}$. Then, using (a1),

$$\begin{aligned}
v_{r,m+1} &\leq Lv_{x,m} + \rho v_{r,m} \\
&\leq LC_0 m^{\lambda_0} + \rho D_0 m^{\lambda_0} \\
&= (LC_0 + \rho D_0)(m+1)^{\lambda_0}(m/(m+1))^{\lambda_0} \\
&\leq (LC_0 + \rho D_0)\sigma(m+1)^{\lambda_0} \\
&\leq D_0(m+1)^{\lambda_0},
\end{aligned}$$

completing the induction step and thus the proof of the lemma.

Proof of Lemma 4.2. By boundedness of ψ, the sequence $\{r_t\}$ generated by (4) is bounded and thus trivially square integrable. Hence, in view of the above lemma A, it suffices to show that G is Lipschitz continuous in x and a contraction mapping in r. As by assumption the first derivative of ψ is uniformly bounded, G is clearly Lipschitz continuous in x with Lipschitz constant $L = M_{\psi}|\bar{C}_r|$. (If A is a matrix, then $|A| := \max\{|Ax| : |x| = 1\}$.) Similarly, let $\nabla_r G$ denote the matrix of partial derivatives of G with respect to r. Note that $|\nabla_r G(x, r, \bar{\theta})|$ is the square root of the maximal singular value of $\nabla_r G$, and thus by a well-known result from linear algebra,

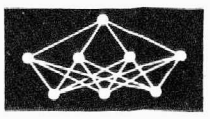
$$\begin{aligned}
|\nabla_r G(x, r, \bar{\theta})| &\leq \left(\text{trace}\left(\nabla_r G(x, r, \bar{\theta})\, \nabla_r G(x, r, \bar{\theta})'\right)\right)^{1/2} \\
&\leq M_{\psi}\left(\text{trace}\left(\bar{D}_r \bar{D}_r'\right)\right)^{1/2} \\
&= M_{\psi}|\,\text{vec}(\bar{D}_r)| \\
&=: \rho.
\end{aligned}$$

By assumption, $\rho < 1$. As clearly,

$$|G(x, r_1, \bar{\theta}) - G(x, r_2, \bar{\theta})| \leq \sup_r |\nabla_r G(x, r, \bar{\theta})|\,|r_1 - r_2| \leq \rho|r_1 - r_2|,$$

G is a contraction mapping in r, thereby completing the proof of lemma 4.2.

Proof of theorem 4.3. We verify then conditions of corollary 3.5 of Kuan & White (1990), which we shall briefly refer to as [KW]. Their conditions A.4 and C.3 are explicitly assumed (our assumptions A.3 and A.4). It follows from lemma 4.2 that $\{r_t\}$ and thus also $\{\xi_t\}$ are bounded sequences NED on $\{V_t\}$ of size -1, where $\xi_t = [x_t', r_t']'$ which establishes condition C.1 of [KW]. Let M_{ξ} be an upper bound for the sequence $\{\xi_t\}$, and let $K_{\xi} := \{\xi : |\xi| \leq M_{\xi}\}$. Condition C.2 of [KW] requires that in $K_{\xi} \times \Theta$, both $F(\xi, \cdot)$ and $\nabla_{\theta} F(\xi, \cdot)$ satisfy a Lipschitz condition with Lipschitz constants $L_1(\xi)$ and $L_2(\xi)$, respectively, where L_1 and L_2 are Lipschitz continuous in ξ, and that both $F(\cdot, \theta)$ and $\nabla_{\theta} F(\cdot, \theta)$ satisfy a Lipschitz condition. It is straightforward to show that continuous differentiability of A.2(i) ensures these Lipschitz conditions. See also corollary 4.1 of Kuan & White (1990).



Proof of theorem 4.4. We verify then conditions of corollary 3.6 of [KW]. Lemma 4.2 ensures that $\{r_t\}$ is NED on $\{V_t\}$ of size -8. Stationarity of $\{x_t\}$ implies that $\{r_t\}$ is also stationary. Hence, $\{\xi_t\}$ is a stationary sequence NED $\{V_t\}$ of size -8, which establishes condition D.1 of [KW]. Condition D.2 of [KW] follows from B.3 and the moment condition of B.1. Finally, as in the preceding proof, four times continuous differentiability of B.2 ensures the Lipschitz conditions imposed in condition D.3 of [KW]. See also corollary 4.2 of Kuan & White (1990).

References

[1] Billingslex, P. (1968): *Convergence of probability measures.* New York: Wiley.

[2] Elman, J. L. (1988): *Finding structure in time.* CLR Report 8801, Center for Research in Language, University of California, San Diego.

[3] Gallant, A. R., White, H. (1988): *A unified theory of estimation and inference for nonlinear dynamic models.* Oxford: Basil Blackwell.

[4] Hornik, K., Kuan, C.-M. (1990): *Convergence analysis of local feature extraction algorithms.* BEBR Discussion Paper 90-1717, College of Commerce, University of Illinois, Urbana-Champaign.

[5] Jordon, M. I. (1985): *The learning of representations for sequential performance.* Ph. D. Dissertation, University of California, San Diego.

[6] Jordon, M. I. (1986): *Serial order: a parallel distributed processing approach.* ICS Report 8604, Institute for Cognitive Science, University of California, San Diego.

[7] Kuan, C.-M. (1989): *Estimation of neural network models.* Ph. D. thesis, Department of Economics, University of California, San Diego.

[8] Kuan, C.-M., Hornik, K., White, H. (1990): *Some convergence results for learning in recurrent neural networks.* Discussion Paper 90-42, Department of Economics, University of California, San Diego.

[9] Kuan, C.-M., White, H. (1990): *Recursive M-estimation, nonlinear regression and neural network learning with dependent observations.* BEBR Working Paper 90-1703, College of Commerce, University of Illinois, Urbana-Champaign.

[10] Kushner, H. J., Clark, D. S. (1978): *Stochastic approximation methods for constrained and unconstrained systems.* New York: Springer Verlag.

[11] Kushner, H. J., Huang, H. (1979): Rates of convergence for stochastic approximation type algorithms. *SIAM Journal of Control and Optimization,* **17,** 607—617.

[12] Ljung, L. (1977): Analysis of recursive stochastic algorithms. *IEEE Transactions on Automatic Control,* **AC-22,** 551—575.

[13] McLeish, D. (1975): A maximal inequality and dependent strong laws. *Annals of Probability,* **3,** 829—839.

[14] Norrod, F. E., O'Neill, M. D., Gat, E. (1987): Feedback-induced sequentiality in neural networks. In *Proceedings of the IEEE First International Conference on Neural Networks* (pp. II: 251—258). San Diego: SOS Printing.

[15] Oja, E. (1982): A simplified neuron model as a principal component analyzer. *Journal of Mathematics and Biology,* **15,** 267—273.

[16] Oja, E., Karhunen, J. (1985): On stochastic approximation of the eigenvectors and the eigenvalues of the expectation of a random matrix. *Journal of Mathematical Analysis and Applications,* **106,** 69—84.

[17] Rumelhart, D. E., Hinton, G. E., Williams, R. J. (1986): Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, & The PDP Research Group, *Parallel distributed processing: Explorations inthe microstructures of cognition,* (pp. I: 318—362). Cambridge, MA: MIT Press.

[18] Sanger, T. D. (1989): Optimal unsupervised learning in a single-layer linear feedforward neural network. *Neural Networks,* **2,** 459—473.

[19] Williams, R. J., Zipser, D. (1988): *A learning algorithm for continually running fully recurrent neural networks.* ICS Report 8805, Institute of Cognitive Science, Unviersity of California, San Diego.

Books Alert

Computational Models of Learning in Simple Neural Systems. Ed. Robert D. Hawkins and Gordon H. Bower. -San Diego, CA: Academic Press, 1989, 321 pp. $ 29.50 softcover, $ 59.50 hardcover.

Computer Systems Performance Management and Capacity Planning. John Cady and Bruce Howarth. —New Jersey, Prentice hall, Englewood Cliffs, 1990, 310 pp., $ 43.20.

Computational Vision. Harry Wechsler. -London, Academic Press 1990. 496 pp.

This book develops an integrated theory of the workings and implementation of computational vision. Using a synergetic approach, the author draws from a broad range of scientific endeavors and attempts to link human and machine vision. The sheer complexity and robustness of the visual task is a unifying theme, and the book facuses on those task characteristics that make vision computationally feasible.

The Emperor's New Mind Concerning Computers, Minds, and The Laws of Physics. Roger Penrose. -Oxford, Oxford University Press 466 pp. ISBN 0-19851973-7. $ 24.95.

Foundations of Neural Networks. Tarun Khanna. -Reading, MA: Addison-Wesley, 1990, 196 pp., paper, $ 26.95, ISBN 0-201-50036-1.

The contents are as follows: Introduction, Associative Memory, The Perceptron, The Delta Rule and Learning by Back-Propagation, Some Learning Paradigms, and the Hopfield and Hoppensteadt Models.

Neural Models of Plasticity. Experimental and Theoretical Approaches. Ed. John H. Byrne, William O. Berry. -London, Academic Press, 1989, 438 pp.

This book explores the role of neuronal plasticity in learning, memory, and complex brain functions by combining theoretical and empirical approaches.

Handbook of Neural Computing Applications. Ed. Cliff Parten, Craig Harston, Alianna Maren and Robert Pap. -London, Academic Press 1990. 400 pp.

Here is a comprehensive guide to help you learn the essential architectures, processes, implementation methods, applications of neural computing systems. Unlike purely theoretical books, this handbook shows how to apply neural processing systems to problems in neurophysiology, control theory, learning theory, pattern recognition, and similar areas. This book not only fully discusses neural network theories, but it also shows you where they originated, how they can be used, and how they can be developed for future applications.

TEACHING STRATEGIES FOR ARTIFICIAL NEURAL NETWORK LEARNING

S. Nordbotten)*

Abstract:

This paper presents an evaluation of the effects of variation of training set size, ordering of examples in the training set, adjustment (learning) rate, and reinforcement on pattern recognition in artificial single layer neural networks (ANNs) which use a learning algorithm based on the Widrow-Hoff principle. These parameters can be considered as alternative teaching strategies for ANNs.

The evaluation has been carried out as a set of simulation experiments on synthetic sets of patterns. The results indicate that for the type of pattern identification considered, learning in ANN is sensitive to the teaching strategy chosen.

1. Teaching and learning

A number of learning algorithms for different artificial neural networks have been developed in recent years [RUMELHART 1988]. The merits of different learning algorithms are currently being studied from theoretical as well as empirical perspectives [NORDBOTTEN 1990b].

In human training systems, there are two actors, the learner and the teacher, and material used for instruction. The complexity of the network topology and learning ability of a learner may be considered as inherited. How effectively a learner learns a given set of examples may depend, however, on the teacher's strategy with respect to the size of the training set used, the order in which the examples are presented, the intensity by which the examples are introduced, and the reinforcement of the teaching material. In the present study we consider the material used for instruction as predetermined.

In the development of an artificial network the designer will frequently define a network which can be trained by examples, choose an adequate learning algorithm suited for the network, and rely on a set of examples which serve as training material. The success of training can be evaluated in several ways:

1) during the training,
2) by evaluation tests or
3) by experience through practical application.

In the study reported in this paper, the impact of several teaching strategy factors has been investigated. The first factor investigated is the size of the training set used. A similar investigation for learning in stochastic knowledge bases used for consultation systems, has been reported in a previous paper [NORDBOTTEN 1991].

The second factor investigated, is the sequence in which the training examples are introduced to the learner. As for a human learner, we assume that the order in which training examples are presented may be important for ANN learning. Learning basic and simple examples before more complex is a frequently applied teaching strategy.

The size of the adjustment rate used by a training algorithm reflects how fast the learner adjusts to an error made. We assume that the adjustment rate can be controlled by the teacher and therefore also belongs to the teaching strategy.

A learner can be exposed repeatedly to each individual example in a training set a number of times before the teacher proceeds to the next example. We will call this strategy concentrated reinforcement. Alternatively, the learner can be exposed to each example of the training set sequentially, and then the exposition for the complete set is repeated a certain number of times. This strategy we will call dispersed reinforcement. Mixed strategies may also be designed by dividing the training set into partitions the examples of which are presented using concentrated reinforcement while the partitions are reinforced in a dispersed manner.

In reinforcement another important factor is the number of repetitive presentations of the training set to the learner. This factor we call the number of reinforcement cycles.

The performance of a trained network in our investigation is considered to be the ability of the network to correctly identify patterns. The overall aim of a teaching strategy is either to give the learner some predetermined performance level using a minimum of teaching investment, a maximum performance level by means of given teaching resources or a maximum of some weighted combination of performance level and teaching investment. We shall return in following

*) Prof. Svein Nordbotten
Department of Information Science
University of Bergen
N-5008 Bergen
NORWAY

sections to the measurement of performance and teaching investment.

2. Teaching strategies

2.1 The Artificial Neural Network

The network model used in this study is a single layer network with M input sources providing simultaneous binary inputs denoted o[i, 0], i = 1 . . M, with value 0 or 1 to a set of N neurons which each generate an output o[j, 1], j = 1 . . N. The vectors of M input elements and N output elements are referred to respectively as a pattern and its identification.

A neuron is a processing unit characterized by its activation level and its output. The activation level is determined by the activation function:

$$a[j, 1] := \text{SUM}[i]\ w[i, j]*o[i, 0], \quad \text{for } j = 1, \ldots, N, i = 1, \ldots M,$$

where w[i, j] denotes the weight between the input source i and the neuron j. The activation level is a real variable.

The output o[j, 0] of the neuron is a binary variable:

$$o[j, 1] := 1 \quad \text{if } a[j, 1] > a[k, 1], \text{ for all } k <> j,$$
$$o[j, 1] := 0 \quad \text{if } a[j, 1] <= a[k, 1] \text{ for one or more } k <> j, \quad k, j = 1, \ldots, N.$$

This implies that a neuron never creates an output vector with more than one non-zero element. Usually an output vector will have one single non-zero element indicating the position of the pattern identification. In some situations, the output vector may have only zero elements, indicating that the network was unable to make an identification. One important aim is to train the ANN recognize the correct pattern identifiers for the input patterns.

In the real world applications we have in mind, there are frequently several different patterns associated with identical target vectors or pattern identifiers. This reflects the possibilty of noise, uncertainty, or errors in the pattern.

2.2. The Learning algorithm

The learning algorithm used is based on the well known Widrow-Hoff algorithm [WIDROW 1960, 1985]. The core of the algorithm used in this investigation consists of two steps:

Step I: The forward computation by which the network, based on current knowledge, computes an activation vector a[j, 1], j = 1 . . N, according to:

```
for i := 1 to M do
    for j := 1 to N do
        a[j, 1] := a[j, 1] + w[i, j]*o[i, 0];
```

and,

Step II. The backwards computation of adjusted weights based on comparison of the target output vector t[j, 1], j = 1 . . N, from the training set and the computed activation vector from Step I.:

```
for i := 1 to M do
    for j := 1 to N do
        w[i, j] := w[i, j] + rate*{o[i, 0]* (t[j, 1] -
        a[j, 1])}.
```

Initially, all weights are set equal to zero. Rate is here the adjustment rate. During the learning process, the algorithm can be repeated in reinforcement cycles. Note that it is the computed activation vector, not the binary output vector, which is compared with the target vector.

The algorithm has the property that it adjusts the weights to minimize the sum of squares of the differences between the elements of the computed activation vector and the target vector. The sum of square errors (SSE) each pattern k of a training set of P patterns is:

```
for k := 1 to P do
for j := 1 to N do
        SSE[k] := SSE[k]+(t[j, 1] - a[j, 1])**2,
```

while the mean square error (MSE) for a training set of P patterns after a learning cycle is:

```
for k := 1 to P do MSE := MSE + SSE[k]/P
```

A MSE decreasing in value from one reinforcement cycle to the next will indicate improved learning.

2.3. Strategy factors

2.3.1 The strategy vector

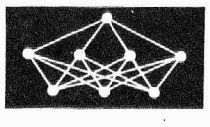
The teaching strategies can be represented in the vector space

$$V = (S, O, L, C, R)$$

where S represents the size of a training set generated randomly from a probability distribution reflecting the patterns of the domain of interest, O represents training set order, L denotes the adjustment rate, R the reinforcement strategy, and C the number of reinforcement cycles.

2.3.2 Size of training set, S

A network's ability to learn a set of different patterns is a main property of the network. In evaluating



a network and its associated learning algorithm, the network can be exposed to and taught pattern sets randomly generated from the domain of interest. After learning, the same set of patterns can be presented to the network to determine how well the network has learned to identify the patterns.

The size of the training set can be easily varied. Two factors must be distinguished. One is the number patterns in the set. Keeping in mind that the training sets are random samples, subsets of identical patterns may exists. The second factor is the number of different target vectors or pattern identifiers present in the set. To each pattern identifier, a subset of different patterns may be associated. The differences are assumed to represent noises acting on the patterns. The ratio between the total number of patterns in the training set and the number of different pattern identifiers therefore indicates the average number of instancies of each identifier within the training set. We would expect that the percent of different patterns correctly identified decreases as the number of patterns increases.

2.3.3 Order of presentation, O

In human teaching, a common strategy is to present to the learner simple patterns prior to more complex patterns. In our investigation a pattern is described as a binary vector with zero and one value elements. In all our patterns, the number of non-zero elements is less than the number of zero elements. The degree of complexity of a pattern is defined by the number of non-zero elements.

Our hypothesis is that learning will be more efficient if the network is exposed to a training set ordered by increasing complexity than if the network is exposed to the same set of patterns presented in a random order.

2.3.4 Adjustment rate, L

The adjustment rate can be considered as the intensity by which the learner is led or instructed to react to errors it makes in recalling a pattern when it learns the correct answer or the target vector. A adjustment rate of >1 corresponds to an over-reaction while a adjustment rate of 0 corresponds to ignorance of errors, or inability to adjust knowledge to facts. We consider adjustment rates in the interval $0 < \text{rate} < 1$ only.

A high rate should be expected to give a fast adjustment to the current pattern. Aplied in a situation with many different patterns it can make harmful disturbance of the weights learned about other patterns. In such a situation, we would expect that a slower adjustment of the knowledge combined with more reinforcement cycles a more safe teaching strategy.

2.3.5 Reinforcement cycles, C

Like human learning, the artificial network's ability to identify a pattern is assumed to improve with reinforcement. We would expect that the ability of a network to correctly recognize a pattern increases with the number of times the pattern has been exposed to the network. However, improvement by repetition would be expected to approach asymptotically a limit, above which no further improvement can be expected.

9.3.6 Reinforcement dispersion, R

A related question is how reinforcement should be organized. We consider two alternative ways in which the reinforcement can be organized, and denote these as dispersed and concentrated reinforcement. In dispersed reinforcement, the different patterns are exposed to the network one by one in some sequence which is repeated in a prescribed number of cycles. In concentrated reinforcement, each pattern is exposed repeationally to the network a prescribed number of times before the next pattern.

With a small number of simple patterns, we expect that concentrated reinforcement would be a wise strategy. However, with an increasing number of patterns, concentrated reinforcement might result in destruction of the knowledge of the first patterns before the last were learned.

3. Experimental design

3.1 Overall design

To study teaching strategies, a set of simulation experiments were carried out. Each experiment consisted of two steps, training and testing. Each training step was designed with a specific training set size, ordering of the patterns, adjustment rate, reinforcement cycle, and reinforcement organization. During this step the mean square error was also carried out. The second step of the experiment was performance tests of the network. In this step each pattern of the training set was recalled one by one to let the ANN compute the output vector, and the percentage of correctly identified patterns in the set was computed.

Two evaluations metrics were thus computed and used:

1) MSE of the training set after learning,
2) PCT of correctly identified patterns in recall from the set.

The mean square error is probably the more general indicator of how the network will work within the domain of interest. Decreasing MSE from one experiment to another indicates improving performance. The percentage of correctly identified patterns is more easily understood and directly interpretable measure. Increasing PCT from one experiment to another indicates superior performance in the second experiment.

3.2 Training sets

The training sets of patterns used were obtained by random generation from a probability distribution [NORDBOTTEN 1990a]. In this distribution, which we assume represent our real world domain of interest, the different target vectors were assigned probabilities, and for each target vector, the element of the pattern vector was assigned a conditional probability. The generation process produced pattern identifiers according to their assumed frequency of occurence. For each generated identifier, an associated pattern was generated with errors or noise included in accordance with the assumed error probability.

The mapping between pattern vectors and target vectors is therefore many-to many. Different input vectors can be associated to the same target vector corresponding to a synonym situation, while different target vectors can be associated to the same input vector corresponding to a homonym problem. The latter is obviously the more serious for pattern identification.

The size of the pattern and target vectors were both 100 elements. The training sets have been used also in several other experiments and are described in detail in other papers [NORDBOTTEN 1989, 1990b, 1991].

Five statistically independent training sets were generated. Each set was replicated and then sorted by increasing complexity.

The resulting 10 training sets were denoted respectively:

R100	S100
R200	S200
R400	S400
R800	S800
R1600	S1600

in which R and S refer to random and sorted, while the numbers indicate the sizes of the the respective training sets.

3.3 Experiments

22 different experiments were carried out. Each experiment was described by the parameters values S<n> | R<n>, L<r>, and D<c> | C<c> as indicated in Fig. 1.

X01:	R100,	LO.1, D15	X12 —	S100,	LO.1, C15
X02:	R200,	LO.1, D15	X13 —	R100,	LO.1, C15
X03:	R400,	LO.1, D15	X14 —	S1600,	LO.1, C15
X04:	R800,	LO.1, D15	X15 —	R1600,	LO.1, C15
X05:	R1600,	LO.1, D15	X16 —	S800,	LO.1, D15
X06:	S100,	LO.1, D15	X17 —	S800,	LO.1, D20
X07:	S1600,	LO.1, D15	X18 —	S800,	LO.1, D25
X08:	R800,	LO.3, D15	X19 —	S800,	LO.1, D30
X09:	R800,	LO.5, D15	X20 —	S800,	LO.1, D35
X10:	R800,	LO.7, D15	X21 —	S800,	LO.1, D01
X11:	R800,	LO.9, D15	X22 —	S800,	LO.1, D02

Figure 1: List of experiments

R<n> denotes a randomly ordered training set of <n> patterns, while S<n> denotes the same training set sorted by increasing complexity. L<r> denotes an experiment in which an adjustment rate <r> was. D or C indicates whether a dispersed or concentrated reinforcement of <c> cycles was applied.

3.4 Implementation

The experiments were programmed in PASCAL and C, and the simulations carried out on an IBM AIX PS/2 and an IBM RS/6000 computer.

4. Discussion

4.1 Restrictions

The results of the 22 experiments carried out, are summarized in table 9.1 to Table 9.5 attached at the end of this paper. The limitation of the experiments must be emphasized and clearly understood.

The topology of the network studied and the learning algorithm applied are only one pair out of many possible which might have been applied for the same purpose. Another pair may have given quite different results.

The training sets used are composed of synthetic patterns each classified in a simple target pattern. Even though the patterns may satisfy certain conditions for representability of a wide class of real world problems, the patterns cannot be claimed in any way to be universally representative. As representations of images, they have a low degree of resolution and complexity.

4.2 Size and ordering of training set

The results of the simulations presented in *Table 1* indicate a clear covariation between the performance indicators and the size of the training set as well as the number of identifiers included in the sets.

The results illustrated in Fig. 2 confirm the expected relations between performance indicators and size of training set. MSE has a increasing value while the PCT on the other hand has a decreasing value when the training set is increased. Bearing in mind the lo-

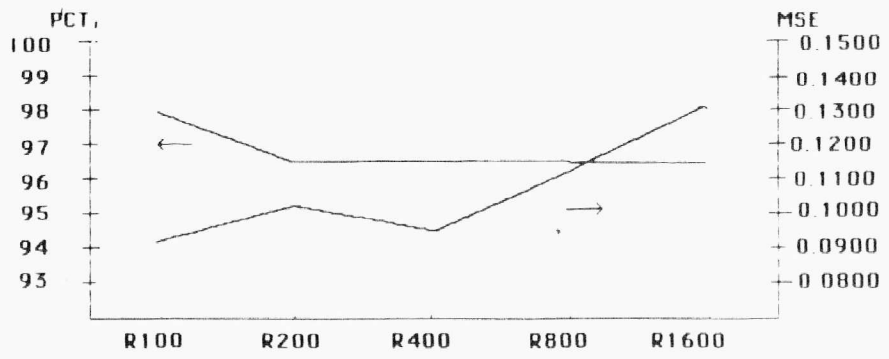


Figure 2: Performance by size of training set. Dispersed training. Adjustment rate 0.1 and 15 reinforcement cycles.

garithmic scale used for presenting the size, the covariation between the two performance indicators and the size of the training set becomes less significant when the set size is increased.

If we study Fig. 3 in which the horizontal axis of the former figure is exchanged with the number of different pattern identifiers in each set from *Table 1*, quite similar results are obtained, but the relations between the performance indicators and the number of identifiers are linear.

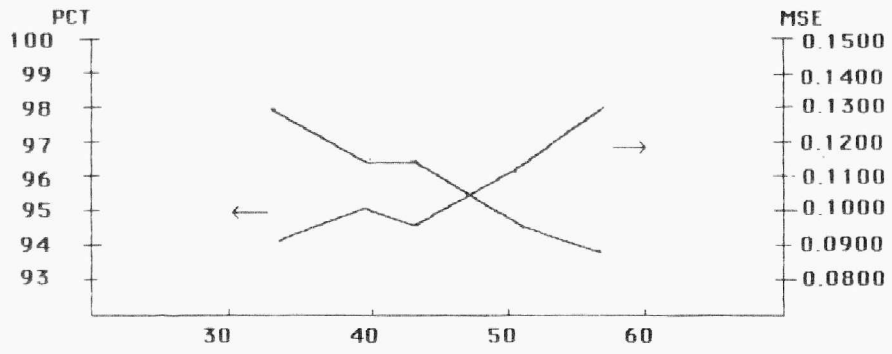


Figure 3: Performance by number of pattern identifiers. Dispersed reinforcement. Adjustment rate 0.1 and 15 reinforcement cycles.

The conclusion we may draw is that the ability of an ANN with a given topology to learn and subsequently make correct identifications within a certain domain decreases by the number of different pattern identifiers existing in the domain. If the number of training patterns is increased, more versions of each patterns are exposed to the network. The network learning will continue, but the improvement will be decreasing and approach a state in which no further improvement can be expected.

Table 2 gives the results of the simulations which were carried out to evaluate the impact of ordering the patterns before they are introduced to the ANN. There is, however, no indication in the results of our experiments that an ordering of the training set has any serious impact on the learning results.

4.3 Adjustment rate

Table 3 presents the results as to the impact of the adjustment rate on the performance indicators. The experiments carried out were based on the training set R800 with Dispersed reinforcemnet in 15 cycles. The relationship between MSE and the value of the adjustment rate is quite clear. The MSE value increases by increasing adjustment rate value as illustrated in Fig. 4. The interpretation must be that in our domain of interest the adjustment rate should be given a small value.

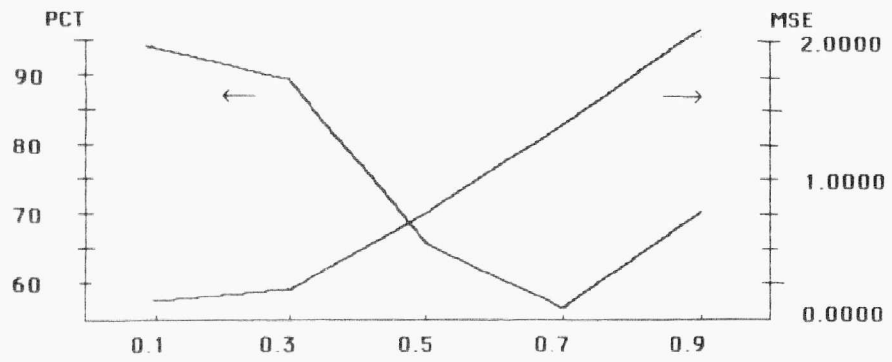


Figure 4: Performance by adjustment rate. Training set R800. Dispersed reinforcement and 15 reinforcement cycles.

The performance expressed by PCT is less obvious. It seems to indicate that the performance curve has a minimum for an adjustment value in the interval 0.5—07. This may, however, be the result of random variations in the training sets. Still, there is no indication that a higher adjustment value should be chosen in preference for a low valued adjustment value. This can be interpreted as support for the assumption that a fast adjustment in a ANN may destroy previously learned knowledge.

4.4 Reinforcement

Table 4 and Fig. 5 give the results of the reinforcement cycle investigation. The experiments were all based on the R800 training set and used an adjustment rate of 0.1. The results support the assumption of significant impact from reinforcement.

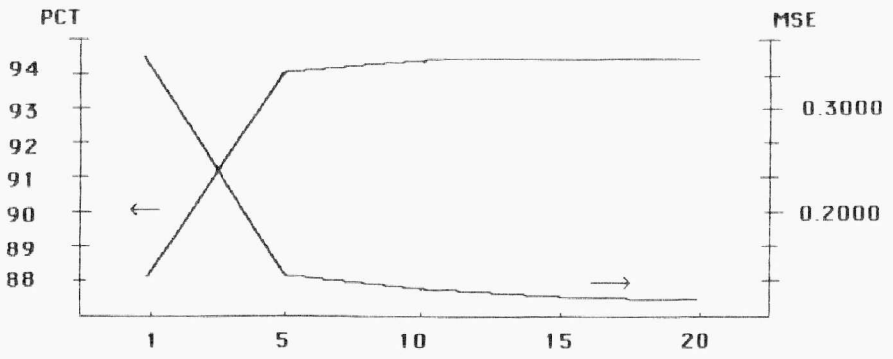


Figure 5: Performance by number of reinforcement cycles. Training set R800. Adjustment rate 0.1. Dispersed reinforcement.

As was expected the performance improved rapidly by number of reinforcement cycles. It is interesting to note that already after 5 cycles the performance obtained was relatively high and further gains in performance by increasing the number of cycles up to 35 were not great.

Table 5 shows the results from experiments with dispersed and concentrated reinforcement in 15 cycles. The experiments were based on the R800 set and the adjustment rate was 0.1. A comparison between the two reinforcement strategies indicates that in all experiments carried out in this investigation, dispersed reinforcement is the superior strategy. Mixed strategies not investigated may, however, give better results than the pure dispersed strategy.

For applications in which the cost of making an additional reinforcement cycle is significant, it is a good reason for considering this cost with the gain in performance.

4.5 Further questions

This paper is one of a series reports on of empirical studies in artificial intelligence and neural networks using simulation. The present investigation was besed on a certain set of assumptions about the functional characteristics of the neurons included. Simulations with non-linear activation functions will also be inves-

tigated and compared with the results presented in the present paper.

As pointed out in the introduction to this section, the practical value of the results from this investigation depends on how representative the patterns we have used are for the applications. If the patterns are not representative, are the results still valid? One way to proceed, which we plan to follow, will be to repeat the present simulations with more complex patterns than used in the present investigation and make comparisons between results from the different investigations.

The experiments and evaluations will also be extended to multi layer networks and learning algorithms for networks with hidden layers of neurons.

5. Conclusions

The investigation carried out indicates that the success of learning in artificial neural networks is sensitive to the teaching strategy applied. In particular, success depends on the number of patterns to be distinguished, the adjustment rate and the number of reinforcement cycles used.

However, the investigation did not give any results supporting the hypothesis that introduction of a sorted sequence of training examples would give better results than a random sequence of examples. Neither did the results support the hypothesis that a concentrated reinforcement would give better learning results than dispersed reinforcement.

Acknowledgement

This work was carried out as part of my research duties at the Univeristy of Bergen. I want to thank my colleagues Associate Professor Joan C. Nordbotten and Research Assistant Atilla E. Gunhan for constructive comments to draft versions of this paper.

References

[1] Nordbotten, S. (1990a): EXPERIMENTOR — An Experimental Environment for Knowledge Based System Simulation, COMPUTATIONAL STATISTICS QUARTERLY, 4, 307—330.

[2] Nordbotten, S. (1990b): Rule Based Systems and Artificial Networks, Invited lecture to be published in TRANSACTIONS OF THE 11th PRAGUE CONFERENCE ON INFORMATION THEORY, STATISTICAL DECISION FUNCTIONS AND RANDOM PROCESSES, August 27.—31, 1990, by the Czechoslovak Academy of Sciences and the Faculty of Mathematics and Physics, and Charles University of Prague.

[3] Nordbotten, S. (1991): Machine Learning of Probabilistic Knowledge Bases, COMOPUTATIONAL STATISTICS QUARTERLY, (to be published).

[4] Widrow, B. and Hoff, M. E. (1960): Adaptive Switching Circuits, 1960 IRE WESCON CONVENTION RECORD, IRE, N. Y., pp. 96—104.

[5] Widrow, B. and Stearn, S. D. (1985): ADAPTIVE SIGNAL PROCESSING, Prentice-Hall, N. Y.

Tables

Set	Number of pattern identifiers	Mean square error	Pct. of correctly identified patterns in recall
R100	33	0.0913	98.0
R200	40	0.1023	96.5
R400	43	0.0949	96.5
R800	51	0.1187	94.5
R1600	57	0.1302	93.9

Table 1: Performance by size of the training set. Adjustment rate 0.1. Dispersed reinforcement with 15 cycles.

Size	Training set order			
	Random		Sorted	
	MSE	Pct. correctly identified patterns by recall	MSE	Pct. correctly identified patterns by recall
100	0.0913	98.0	0.0619	97.0
1600	0.1302	93.9	0.1303	93.2

Table 2: Performance by training set size and ordering. Adjustment rate 0.1. Dispersed reinforcement with 15 cycles.

Adjustment rate	Mean squared error	Pct. correctly identified patterns in recall
0.1	0.1187	94.5
0.3	0.2145	89.0
0.5	0.7478	66.0
0.7	1.4077	56.5
0.9	2.0943	65.1

Table 3: Performance by adjustment rate for set R800. Dispersed reinforcement with 15 cycles.

Cycles	Mean square error	Pct. of correctly identified patterns in recall
01	0.3577	88.1
05	0.1392	94.0
10	0.1235	94.3
15	0.1187	94.3
20	0.1164	94.3
25	0.1152	94.3
30	0.1145	94.3
35	0.1141	94.4

Table 4: Performance by reinforcement cycles for set R800. Adjustment rate 0.1. Dispersed reinforcement.

Set	Pct. of correctly identified patterns in recall	
	Dispersed reinforcement	Concentrated reinforcement
R100	98.0	94.3
S100	97.0	83.0
R1600	93.9	85.8
S1600	93.2	82.0

Table 5: Performance by reinforcement concentration. Pct. of correctly identified patterns. Adjustment rate 0.1. 15 reinforcement cycles.

SELF-REPRODUCIBLE NETWORKS: CLASSIFICATION, ANTAGONISTIC RULES AND GENERALIZATION

Ezhov A. A., Khromov, A. G., Knizhnikova L. A.), Vvedensky V. L.**)*

Introduction

Self-reproducible neural networks (SRN) with synchronously changing neuron thresholds are interesting objects for theoretical investigations and computer modeling [1]. The properties of these networks may have direct biological analogies, so it is natural to treat such objects with minimal restrictions. On the other hand generalization of the model may simplify its investigation and lead to strong theorems. In this communication we present our recent results that may help in formulation of the theory. In particular, we describe the networks with anti-Hebbian bonds which have some interesting properties and naturally lead us to a generalization of the conception of self-reproducibility.

The fundamental model [1] implies an ensemble of neural networks with the same number of neurons N, where networks can exchange information with each other. In the simplest case the fundamental model considers Hopfield networks [2] with the following properties. Every neuron can be in passive (0) or active (1) state. The running state of a network is described by a vector with components V_i = (0 or 1), where $i = 1, \ldots, N$. Synaptic connections between neurons i and j constitute matrix T_{ij} with positive elements for excitatory and negative ones for inhibitory synapses. For any initial state of the network its evolutions is determined in the following way. Action of the rest of the network on the k-th neuron is calculated as

$$F_k = \sum_{i=1}^{N} T_{ki} V_i .$$

If F_k exceeds the threshold U_k for the neuron, it switches into $V_k = 1$ state (it fires), if $F < U_k$ then $V_k = 0$ (stays silent), if $F = U_k$. Later we'll show that monotony is not necessary for self-reproducibility in general, but for the networks with Hebbian connections we have no examples of nonmonotonic informational sets of patterns. The case of a monotonic Hebbian SRN seems to be general enough to obtain an estimate for a cardinal number of SRN sets. We consider just this case.

*) Ezhov A. A., Khromov A. G., Knizhnikova L. A.
Affiliated Branch of Kurchatov Institute of Atomic Energy, 142092, Troitsk, Moscow Region, USSR.
**) Vvedensky V. L.
Kurchatov Institute of Atomic Energy, 123182, Moscow, USSR.

1. Hebbian SRN with Monotonic Informational Set of Patterns

First we consider necessary and sufficient conditions for self-reproducibility in the monotonic case. Suppose that the learning rule is Hebbian [2]

$$T_{ij} = \sum_{s=1}^{M} (2V_i^s - 1)(2V_j^s - 1), \quad i, j = 1, \ldots, N; \quad T_{ii} = 0;$$

$$s = 1, \ldots, M$$

We introduce an m-basis for the set $\{V^1, \ldots, V^N\}$, which consists of $L = N + 1$ vectors B^l, $l = 1, \ldots, L$ [3]. It is sufficient to form the matrix $J_{si} = V_i^s$ and to find all identical columns of the matrix to generate this basis. $B_i^l = 1$ at the positions of the columns of l-th type and equals 0 everywhere else. All the vectors V^s and all the network's attractors can be presented as a combination of these basic vectors (see *Fig. 1*).

	B^1	B^2	B^{N-1}	B^N	B^{N+1}
V^1	11111111111	111111	111	111	000000
V^2	11111111111	111111	111	000	000000
. . . . .					
V^{N-1}	11111111111	111111	000	000	000000
V^N	11111111111	000000	000	000	000000

Figure 1. Informational set of patterns. Groups of neurons corresponding to different basic vectors (for which $B_i^l = 1$) are shown

The neuron i belongs to the basic vector B^l if $B_i^l = 1$. In our case the Hebbian matrix of connections T_{ij} has such a block structure, such then V_k remains unchanged. Matrix T_{ij} is symmetrical with zero diagonal elements. Evolution of such a network ends up in a stationary state corresponding to the minimum of energy functional [2].

We introduce a new important feature — that is a mechanism of synchronous change of all neuronal thresholds in every network of the ensemble. We assume that the thresholds for all neurons in a network are equal and can be changed in synchrony between

U_{min} and U_{max}. For any matrix of connections T_{ij} one can choose U_{min} so that the network becomes "epileptic", that is $V_i = 1$ for every neuron. We suppose that this extremal state initiates an information transfer between this "donor" network and other "acceptor" network(s). In a similar way U_{max} can be chosen so that the network falls into "coma", that is $V_i = 0$ for all neurons and this state inactivates information channels. In the following, U_{min} and U_{max} are chosen so that during the sweep of the threshold in this range the network transits from "epilepsy" to "coma", sending information outside in the form of quasi-stationary patterns of activity.

We consider variations of the threshold (U) to be adiabatic, so that at any U the network has time to relax to a certain stationary state. It remains stationary until U reaches a new value when the network relaxes to a new state. In such a way during the sweep between U_{min} and U_{max} the neural network passes finite number of stationary (quasi-stationary) states which can be considered as an "informational set" transmitted into an acceptor network. The extremal "epileptic" and "coma" states are not included in this set. The transfer of the informational set leads to modification of synaptic connections in the acceptor network in accordance with (we consider this to be simplest case), the Hebbian rule [2].

Let us see what would happen in a linear chain of the neural networks of that kind (see *Fig. 2*). Suppose, that only the first network ($k = 1$) is trained, that is having a nonzero matrix of connections. All the rest ($k > 1$) are ignorant and have zero matrixes — "tabula rasa". We assume that the information transfer is unidirectional to the next neighbor — from network k to $k + 1$. In the beginning the sweep of the threshold in all the networks produces information transfer only from the first one which trains the second network. The second network trains the next one during the next sweep of the threshold and so on. The repetition of information transfer from network k to $k+1$ does not change the set of attractors in the former. One observes propagation of the "learning wave" in the chain.

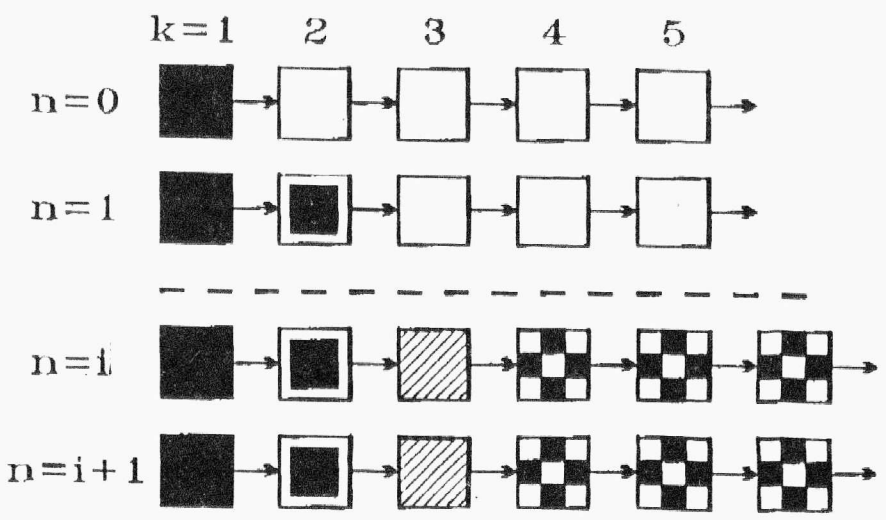


Figure 2. Transformation in the chain of neural networks transmitting information to nearest neighbour (k to k + 1). in the course of repetitive sweep of the neural threshold (n — number of sweeps). Different symbols means different information sets in particular networks. Initially, only the first network contains a nonzero information set — solid black; the others are empty — open boxes. After i sweeps a row of self-reproducible networks appears — checkerboards. Networks with transitional information sets are also present — other symbols.

The remarkable feature of this process, observed in computer simulations [1], is the emergence of identical networks after a small number of transitional networks in the beginning of the chain. The population of the identical networks grows with every sweep of the threshold; they transmit the same set of patterns to the neighbor and build the same matrix of connections. In fact these networks are copying themselves and we call them self-reproducible networks. The simplest version of the model may be generalized in different ways considering other learning rules, nonsymmetrical matrices of connections (and more general forms of network attractors), other geometries (two-dimensional, for cxample) of network ensembles, etc. [1], though the phenomenon of developement of self-reproducible networks seems to be independent of all these complications! We mentioned already the need to consider observed phenomenon in the simplest form in order to find the fundamental results.

Therefore, we discuss the described chain of Hopfield networks with Hebbian interconnections. Despite the fact that this simplest scheme gives random (in some sense) SRN it is useful because it gives some experience and knowledge on the structure of these interesting objects. One observation is that for the networks with Hebbian connections, an informational set of patterns of SRN is monotonic (we didn't find any counter example so far). It means that in a set $\{V^1, \ldots, V^N\}$, ordered in accordance with the threshold value at which the patterns arise, there are no two patterns V^k, V^m, such that $k < m$, but for some i $V_i^k < V_i^m$. In a monotonic set under suitable permutation of neuron indexes all patterns can be presented in such a manner that $V_i = 1$ for $i = < N_k$ and $V_i = 0$ for $i > N_k$ (see *Fig. 3a* and compare with *3b*).

	a)		b)
V^1	11111111111111110000	V^1	11111111110000000000
V^2	11111111111100000000	V^2	11111111111111100000
V^3	11111111110000000000	V^3	11111111111100000000
V^4	11111000000000000000	V^4	11111111100000000000

Figure 3. Examples of pattern sets: a) monotonic, b) nonmonotonic.

that for neurons i and j belonging to basic vectors B^l and B^k respectively, their synaptic junction efficiency equals

$$T_{i(l)j(k)} = n - 2|l - k|, \quad T_{ii} = 0. \tag{2}$$

If the network is initially in epileptic state $V^0 = (1, 1, \ldots, 1)$ then the force's values for neurons of different groups will be equal to

$$F^1 = 2\sigma - n(N + 1)$$

$$F^l = 2\sigma - n(N + 1) + 2\{(l - 1) N - 2a_n \ldots - 2a_{n-l+2}\}, \quad l \geq 2. \tag{3}$$

where $\sigma = \sum_{s=1}^{n} a^s$, $a^s = \sum_{i=1}^{n} V_i^s$

-activity of s-th pattern.

If a network under consideration is self-reproducible, then the minimal force value is for $l = n + 1$. Force value depends on l so that the minimum may be reached only for $l = 1$ or $l = n + 1$ and the latter case takes place if $\sigma > nN/2$.

Then, for threshold $U^0 = F^{n+1}$ the neurons of the $n + 1$-th group (i.e. belonging to basic vector B^{n-1}) may pass into zero state. But this does not mean that networks will pass into the stationary state V^1, because forces acting on the neurons from other groups after the switch off of neurons from the $n + 1$-th group may become smaller then U. Let us introduce a restriction and suppose that for the set of patterns $\{V^1, \ldots, V^n\}$ this is impossible.

More concretely, let us suppose that switch off of the neurons from the k-th group in any state V^{n+1-k} (for the threshold value U^{n+1-k}, for which this state become unstable) changes the actions on the neurons from other groups in such a manner that their new values will be still less then U^{n+1-k} for passive neurons and still more than this threshold for active neurons. In other words let us demand that during relaxation from state V^{n+1-k} to V^{n+2-k} only neurons of the k-th group may change their states. We shall refer to such self-reproducible networks as monotonic with simple dynamics. After simple algebraic transformations one can get necessary and sufficient conditions for realization of such a type of dynamics. They have the following form

$$\begin{aligned} &\gamma N < a^1 < N, \\ &\gamma a^1 < a^2 < a^1, \\ &\ldots\ldots\ldots\ldots\ldots \\ &\gamma a^{n-1} < a^n < a^{n-1}, \end{aligned} \tag{4}$$

where $\gamma = n/(n + 2)$,

$$a^1 + a^2 + \ldots + a^n > nN/2.$$

We may calculate under these conditions the number of classes of SRN with monotonic informational sets and simple dynamics. Each of this class includes pattern sets which differ only by all possible permutations of neurons.

It can be shown that the number of classes of SRN with fixed number of patterns in informational set $K(n, N)$ is less then

$$K_{max}(n, N) = C(n)N^n, \tag{5}$$

where

$$C(n) = (1 - \gamma)(1 - \gamma^2)\ldots(1 - \gamma^n)/n! \tag{6}$$

The actual number of such SRN is sufficiently close to this estimate for large N and N/n (see *Table 1*).

		$K(n, N) \leq K_{max}(n, N)$	
$N = 50$	$n = 2$	384	468
	$n = 3$	3 532	4 181
	$n = 4$	21 820	27 232
$N = 100$	$n = 2$	1 601	1 875
	$n = 3$	30 379	33 450

Table 1. Number of classes of monotonic SRN with simple dynamics

It is also interesting that qualitative properties of SRN depend on how close the pattern activities are to the lower or upper limits of inequalities (4). If $a^k \simeq a^{k-1}$ ($a^0 = N$), then for each threshold value, the network has as a rule only one stationary state and looks like a "generalizing" one. On the other hand, if $a^k \cong \gamma a^{k-1}$ then for each threshold value there exists as a rule many stationary states and the network looks like a "memorizing" one (see *Fig. 4*). (It should be noted that such attractors may have some interesting interpretations [4].

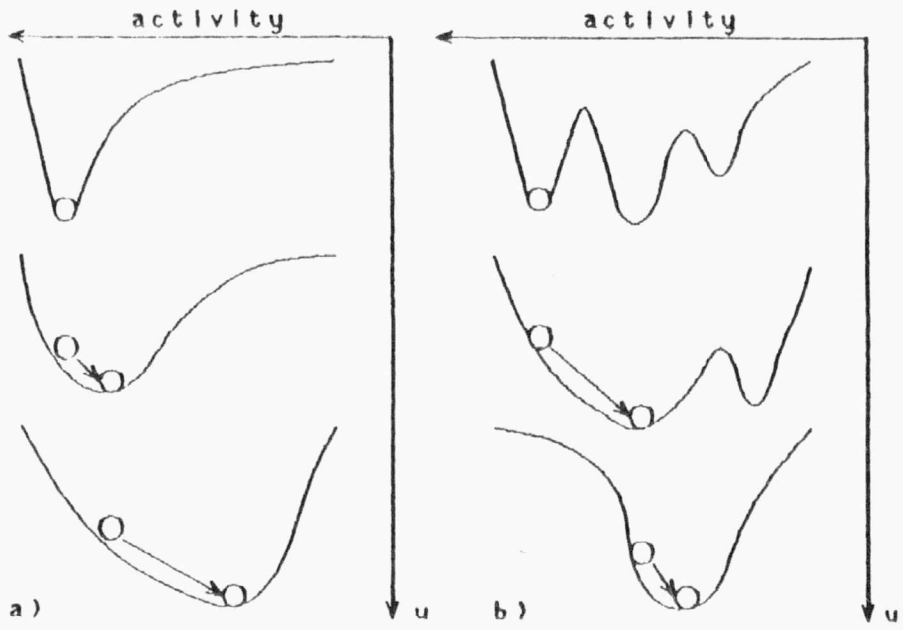


Figure 4. Landscape metaphor for SRN of two different types:
a) "generalizing network has single attractor for all threshold's values U;
b) "memorizing" network has many attractors for different values of threshold.

Are monotony and simplicity of dynamics crucial for self-reproducibility of Hebbian networks? We don't know, but our computer modeling did not give any example of Hebbian SRN having complex dynamics. Is it possible to prove that monotony and simplicity of dynamics are the necessary and sufficient conditions of self-reproducibility on other cases? Probably this task is difficult enough if we shall not limit ourselves to the above case. Indeed, if we consider for example neurons with spin states $\mu_i = \pm 1$ then the emergence of SRN in some cases will seem more complex. It is not difficult to see that for spin model formation of SRN does not mean in general the stabilization and repetition of an informational set of patterns. For example in a 4-neuron self-reproducible network with 3 learned patterns

$\mu_1 = (1, 1, 1, -1)$, $\mu_2 = (1, 1, -1, -1)$,
$\mu_3 = (1, -1, -1, -1)$

the informational set may also consist of the following patterns

$\mu_1 = (-1, 1, 1, 1)$, $\mu_2 = (-1, -1, 1, 1)$,
$\mu_3 = (-1, -1, -1, 1)$.

Hebbian matrices constructed from these two sets are identical.

The cause of nonuniqueness of informational set is the asynchronous random dynamics of Hopfield's networks.

Hence, we must consider self-reproducibility at least as repetition of matrix structure, but not necessarily a repetition of informational sets of patterns. This introduces some complications and makes the problem more general. Is the problem in this new form general enough to make its solution relatively simple?

We think that it is possible to generalize the phenomenon of self-reproducibility further, but first we describe a new method of generation of SRN.

3. Networks with Antagonistic Learning Rules

In contrast to the Hebbian rule (1) we may introduce the so called "anti-Hebbian" learning rule which differs only by the negative sign before sum in (1). This learning rule tends to form "hills", not "holes", near corresponding states V^s; $s = 1, \ldots, n$ (see *Fig. 5*). At first glance the anti- Hebbian rule gives us an example of a law which does not lead to formation of any self-reproducible network. Hence, our observations give the evidence of non-triviality of self-reproducibility. Moreover, for a small number of recorded patterns, n anti- Hebbian networks look like *generators* of SRN with a *direct* Hebbian learning rule.

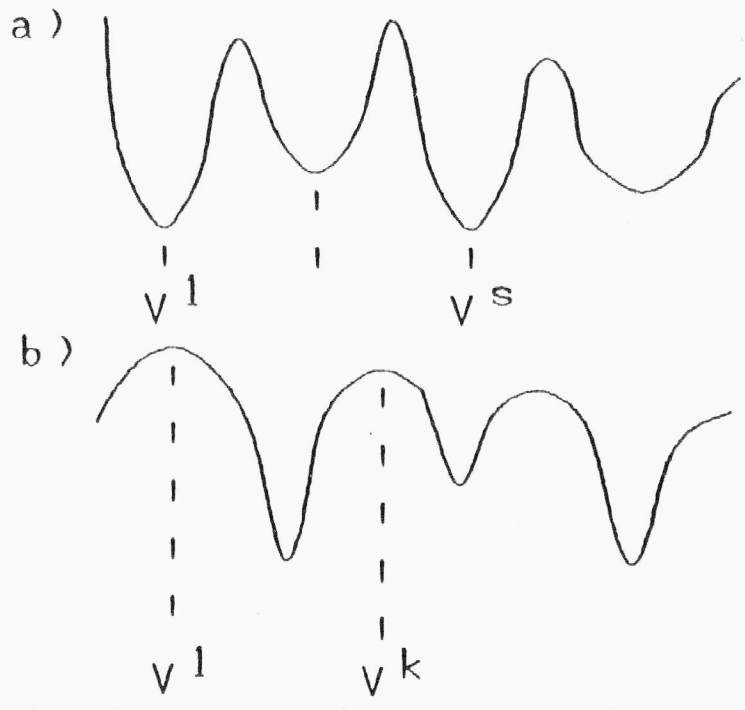


Figure 5. Landscape metaphors for networks with Hebbian (a) and anti-Hebbian learning rules. The use of an anti-Hebbian rule leads to formation of hills near the stored patterns V^s.

For example, if $n = 1$ then after only one threshold sweep the anti- Hebbian network generates informational set of patterns consisting of $Abs(|V^1| - |N - V^1|)$ patterns (V^1-pattern recorded in the anti-Hebbian network) with activities $a^1 = N - 1$, $a^2 = N - 2, \ldots, a^m = N - m$. If $m < 2/3N$ then this set determines a self-reproducible Hebbian network.

This informational set consists of patterns which may be produced by sequential one-neuron switch off.

If two patterns V^1 and V^2 are stored in synaptic bonds of anti Hebbian matrix then one threshold sweep in this network leads to generation of a set consisting of m paterns with activities

$a^1 = N - 1, a^2 = N - 2, \ldots, a^{m-c} = N - (m - c)$;
$a^{m-c+1} = N - (m - c + 2), \ldots,$

$a^m = N - (m - c + 2(m - c))$ where $m = \max(\delta B^1, \delta B^2)$,
$c = \min(\delta B^1, \delta B^2)$,

$\delta B^k = Abs(|B^k| - |\tilde{B}^k|)$, $k = 1, 2$ and B^k, $\tilde{B}^k$ — vectors of m-basis that correspond to mirror-symmetrical columns in matrix V_i^s [3].

If
$m \le N/3$, $c \le m$ or $N/3 \le m \le 2N/3$, $c \le 2N/3 - m + 1$
then this set determines a self-reproducible Hebbian network. It is clear that the informational set of this network consists of patterns which are produced by sequential switch off of one or two neurons.

An Anti-Hebbian network built with $n > 2$ patterns has set of patterns which can be produced by sequential passivizetion of more then two neurons in a step.

Hence, in raising the threshold of the antagonistic network we may obtain an informational set of patterns for a self-reproducible network with a *direct* learning rule. But it is possible that we obtain a monotonic set of patterns, which only sometimes may determine a self-reproducible network and an, antagonistic network is the generator of networks with monotonic learned patterns rather than a generator of SRN.

We do not know exactly the relation between monotony and self-reproducibility. However, the antagonistic networks can give us a quite different view on the phenomenon of self-reproducibility and these networks can be more important than simple generators of a monotonic set of patterns.

4. Generalized Self-Reproducibility

Remember that at first glance, the anti-Hebbian rule looks like an example of a law for which the phenomenon of self-reproducibility does not take place. Indeed, for every set of stored patterns an anti-Hebbian network has attractors quite different from these to the patterns (this is due to formation of "hills" not "holes" near states V^s in a configuration space).

Therefore recording of patterns from an informational set of donor network produces acceptor network with apparently different connections. In a line-

ar chain of networks we shall see the emergence of random-like neural networks. One can think that the "theorem of convergence" for a linear chain of networks must use conditions that exclude the case of the anti-Hebbian law. But one can ask whether an anti-Hebbian network analogous to a self-reproducible (Hebbian, for example) exists. Of course, such analogy can differ substantially from the usual SRN, but can also have something in common with the SRN.

It is surprising enough that such an anti-Hebbian network with a quasi self-reproducible property is not difficult to find. In every chain of such networks permanently appear remarkable areas of networks transferring an equal number of patterns to the next neighbor. The informational sets of such networks differ one from another only by the permutation of neuron indexes. In other words these quasi self-reproducible networks may reproduce themselves with high probability in some generalized sense – the offsprings can differ from the parent by some trivial transformation. Virtually such a network may produce another quasi transient network (with a slightly different set of pattenrs as a rule) but very quickly the area of one-type networks emerges again (see *Fig. 6*).

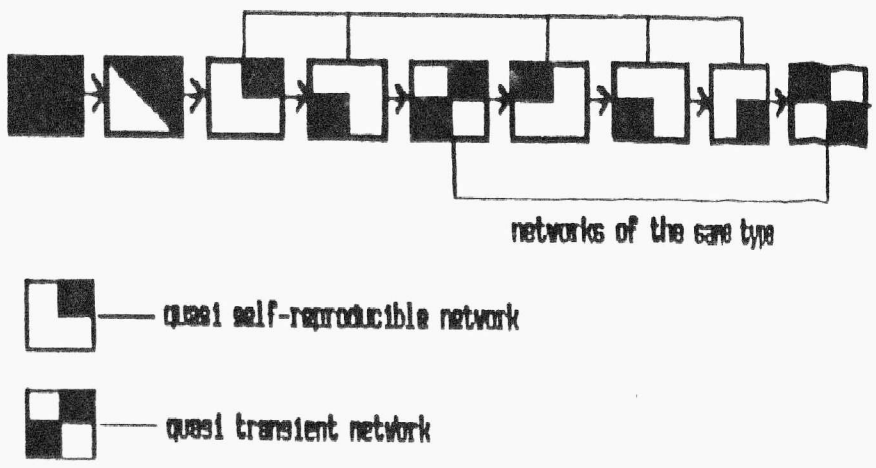


Figure 6. Emergence of quasi self-reproducible and quasi transient neural networks in a linear chain of anti-Hebbian neural ensembles emitting information unidirectionally to the next neighbor. Networks belong to the same type if they differ by a purmutation of neuron's indices only (corresponding boxes are simply rotated).

Networks from such an area are characterized by informational sets which in some sense generate Boolean functions of the same type [5]. Therefore, we may consider the reference phenomenon as reproducibility of network's type. In the general case the initial donor network with anti-Hebbian connections does not produce a final *steady state*- ordinary self-reproducible network in a chain of networks. Nevertheless synchronous repetitive change of threshold brings the evolution of the network's structure to the final *strange attractor* which includes a number of different types of networks. *Fig. 7* presents the structures of such attractors for a small number of neurons.

Computer simulations show that for different numbers of neurons (at least for small ones) and for almost any initial matrix of interconnections of initial network only one such attractor exists.

Degenerate initial matrices can also exist which

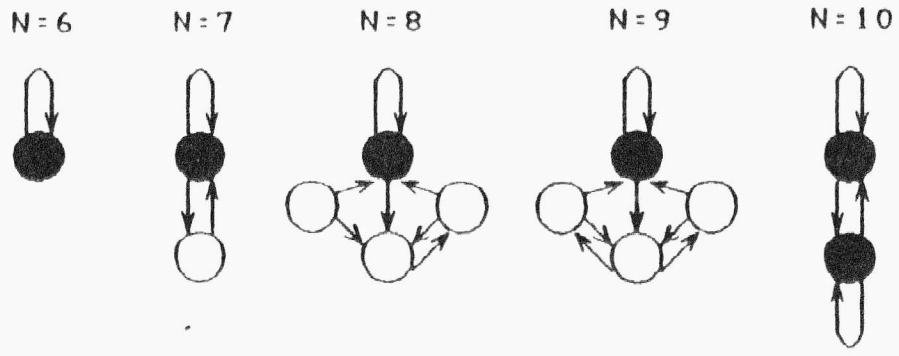


Figure 7. Structures of strange attractors in a space of network configuration types for different numbers of neurons N. Black circles denote classes of quasi self-reproducible neural networks of the same type; white — classes of quasi transient networks of the same type. Arrows show possible transitions between classes in a chain of interacting networks.

lead to formation of so-called *black hole* networks with an empty informational set. For example, an anti-Hebbian network containing one pattern with equal numbers of units and zeros into anti- Hebbian networks produces a *black hole* daughter network. (In the case of a 4-neuron network any initial network gives a final network with an empty informational set — this case is unique). This is similar to famous Honon strange attractor, where some initial data led to the infinite growth of the solution and others led to a strange attractor [6]. The networks appearing during the transition of the system into the strange attractor area may be roughly divided into quasi self-reproducible and quasi transient. The former has a nonzero probability to reproducing a network of the same type; for the latter this probability is vanishing.

Another important feature of these networks is that they may have a nonmonotonic informational set.

Let us summarize the main features of generalized self-reproducibility.

1. Transfer of stationary patterns which are passed by a network with anti-Hebbian connections during a synchronous sweep of the threshold in a chain of identical networks leads to formation of a limited class of networks.
2. These networks can be divided into groups of networks of the same type. The networks of the same type have informational sets, containing patterns differing from each other by permutations of neuron's indices only.
3. *Quasi self-reproducible* networks may reproduce networks of the same type. *Quasi-transient* networks may reproduce only networks of another type.
4. If we define a mapping of the set of all possible networks onto itself taking into account that synchronous changing of thresholds and informational transmission determines the corresponding network to network transformation; then generalized self-reproducibility may be represented as the chaotic moving of trajectory of the point representing the network's structure within the limited area of configurational space.

In conclusion we remark that anti-Hebbian networks show the complexity of the phenomenon of

self-reproducibility in general. The investigation of this phenomenon in its general formulation is related to the study of the chaotic behavior of dynamical systems as a whole and of neural networks in particular [7].

References

[1] A. A. Ezhov, V. L. Vvedensky, A. G. Khromov, L. A. Knizhnikova: Self-Reproducible neural networks with synchronously changing neuron thresholds. — In: "Neurocomputers and attention", Extended Abstracts of International Workshop, Moscow, 1989, p. 102—103.

[2] J. J. Hopfield: Neural networks and physical systems with emergent collective computational abilities. Proceedings of National Academy of Science, USA, 1982, vol. 79, p. 2554—2558.

[3] A. A. Vedenov, A. A. Ezhov, A. M. Kamchatnov, L. A. Knizhnikova, E. G. Levchenko: A study of Hopfleld's model of associative memory. Kurchatov Institute of Atomic Energy, Preprint 4262/1, 1986.

[4] O. Kufudaki, J. Hořejš: Associative neural networks with threshold-controlled attention. In: "Neurocomputers and attention" Extended abstracts of International workshop, Moscow, 1989, p. 130.

[5] A. A. Vedenov, A. A. Ezhov, L. A. Knizhnikova, E. B. Levchenko: Spurious memory in model neural networks. Kurchatov Institute of Atomic Energy, Preprint 4395/1, 1987 (in Russian).

[6] M. Henon: A two-dimensional mapping with a strange attractor, Communications in Mathematical Physics., vol. 50, p. 69, 1976.

[7] K. Kurten: Critical phenomena in model neural networks, Physical Review Letters. A, 1988, v. 129, n3, p. 157—160.

Book Review:

Hecht-Nielsen, R. : Neurocomputing
Addison-Wesley Publishing Co, 1989, pp. 432, ISBN 0-201-09355-3

The book comprises 9 chapters and an Appendix:

1. Introduction: What is neurocomputing?
2. Neural Network Concepts, Definitions and Building Blocks
3. Learning Laws: Self-Adaptation Equations
4. Associative Networks: Data Transformation Structures
5. Mapping Networks: Multilayer Data Transformation Structures
6. Spatiotemporal, Stochastic and Hierarchical Networks: Frontiers of Neurocomputing
7. Neurosoftware: Description of Neural Network Structures
8. Neurocomputers: Machines for Implementing Neural Networks
9. Neurocomputing Applications: Sensor Processing, Conrol and Data Analysis

A. Neurocomputing Projects: Developing New Capabilities that Succeed in the Marketplace

As the contents indicates, the author tries first to classify various concepts (including implementation, management and project planning) and only within this framework he introduces most usual paradigms (of which e. g. backpropagation is handled more consistently using sun — planets metaphor). Generally, this book is more technically oriented (practically omitting neurophysiological motivations together with models serving primarily brain research, like works of Grossberg, Krjukov and others), requires more prerequisites (including some higher mathematics, like calculus, linear algebra, statistics, filters etc) and is more original than Wasserman's introduction [see the page 38]. It also brings exercises, more complete bibliography and examples of real applications.

The approach brings a lot of valuable insights, comments on various experiences and gives theoretical mathematical support wherever possible.

There may still be some objections concerning both organization of the book, style of presentation and incompleteness. Thus e. g. the generality of introductory parts describes too many concepts, used in full generality afterwards only occasionally (slabs, fascicles, classes), which may disgust the reader keen to meet first more concrete facts, and ready to wait for the general architecture description as the need arises. Treatment of neurosoftware is not too instructive: every careful reader will perhaps design the software procedures along similar lines and details of AXON example is not much more than simple exercise in C. Chapter 9 should be better incorporated into the text after chapter 6, some applications being mentioned at least implicitly in the preceding parts anyway. Also section 6. 3 is ordered somehow illogically (neocognitron — combinatorial hypercompression — attentional mechanism: back to neocognitron). The exposition is sometimes not well-balanced, easier parts being often discussed in more detail than the more complicated ones; e. g. in sect. 9. 2, where character recognition seems to be understandable at first sight while "logons" would deserve more space. Of often cited paradigms, I miss ART, "neurons" with decay factor and similar questions, no matter that they are more biologically inspired.

The mentioned objection are however overcame by the positive features of the book, which brings a handful of ideas both from the theory and real-world practice. I can strongly recommend reading it carefully by anyone who wants to make a second step in learning the fascinating yet down-to earth story of neurocomputing.

J. Hořejš

Books Alert

Advances in Neural Information Processing Systems 2 Ed. David S. Touretsky. -San Mateo, CA: Morgan Kaufmann, 1990, 853 pp., bound, $ 35.95, ISBN 1-55860-100-7.

This volume contains the collected papers of the 1989 IEEE Conference on Neural Information Processing Systems-Natural and Synthetic. This collection of over 120 papers represents the increasing cross-fertilization of the interdisciplinary nature of neural network research.

Advanced Neural Computers. Ed. R. Eckmiller. -Amsterdam, Elsevier Science Publishers, 1990, 500 pp. ISBN: 0-444-88400-9.

This book is the outcome of the International Symposium on Neural Networks for Sensory and Motor Systems held in March 1990 in the FRG. The NSMS symposium assembled 45 invited experts from Europe, America and Japan representing the fields Neuroscience.

STATISTICAL MODELS OF INTRANEURAL TOPOGRAPHY

O. Frank)*

Summary

Segregation by modality of various fibers in human sensory nerve fascicles was studied by recordings from thin needle electrodes. In order to analyze data from such experiments, a statistical model was developed to test intraneural modality clustering. This model and some test statistics are presented here.

Introduction

Sensory nerve fascicles in the human arm are composed of bundles of fibers of different modalities. A special electrode for percutaneous recording of fiber activity has been developed and tested at the Huddinge University Hospital in Stockholm. Neurological findings in this research are reported in a series of articles; see, for instance, Hallin, Wiesenfeld-Hallin and Duranti (1986), Hallin (1990) and Hallin, Ekedahl and Frank (1990). The last reference contains a stochastic model of nerve impulse recording that was used in the statistical analysis of the experimental data. The purpose of the present note is to describe this model in a general setting and indicate some extensions of it. The emphasis here is on the probabilistic and statistical aspects of the model.

An Intraneural Recording Technique

Activity in nerve fibres was recorded by using thin needle electrode with a „recording window" at the top. This window is of oval shape and of approximate length 8 and width 3 measured in units of nerve fiber diameters (about 12 μ). Recordings are obtained from a sequence of sites when the needle is moved perpendicular to the nerve. Displacements that are smaller than the length of the window are causing overlaps between neighboring sites which implies that the same fibers might be contributing to several records.
Along the fibers there are nodes at a distance of about 50 μ. These nodes are centers of nerve activity that have to be within the window in order to be contributing to the record. The window's width is a fraction of $3/50 = 0,06$ μ of the internodal distance, and this fraction can be considered as a probability of finding activity in an individual fiber. Now each recorded activity cannot be traced to an individual fiber but only to the set of fibers in the window. With 8 fibers in the window, the probability of at least one recordable activity is $1 - 0,94^8 = 0,39$. This probability reflects the difficulties encountered in trying to find an initial site with activity. After such an initial site has been found, further recordings are obtained from a sequence of sites having about 3 units displacement and 5 units overlap with the previous site. This overlap implies that the recordings from different sites are dependent, and this has to be taken into account when experimental data are used to draw conclusions about the arrangement of fibers of different modalities in the nerve fascicles.
There are four specific modalities of fibers that can be identified by different signals. These modalities are present in the proportions 0. 40, 0. 15, 0. 25 and 0. 20 so that, for instance, the window is expected to have 3.2 fibers of the first modality. Intraneural topography tries to describe how the nerve fascicles are composed of fibers of specific modalities. Are the different modalities randomly mixed or are there any tendencies for fibers of similar modality to be close together in the fascicles? Experimental evidence seems to imply that there is a clustering tendency by modality. The statistical problem is essentially to determine whether data from overlapping sites give significant for clustering.

A Stochastic Model

Let X be the recorded state of a fibre. With k different modalities there are $k+1$ states labeled by 0 for no activity and labeled by 1, . . . , k for the different modalities. The probabilities of different states are denoted by $p_o, \ldots, p_k$. The initial site consists of m fibers of states $X_1, \ldots, X_m$. The next site is obtained by a displacement across h fibers, where h is an integer between 1 and m. This site has m-h fibers in common with the initial site, and the states of its fibres are $X_{1+h}, \ldots, X_{m+h}$. Continuing to label the states in this way, the next site will have the states $X_{1+2h}, \ldots, X_{m+2h}$, and so forth. Let $Y_{01}, \ldots, Y_{k1}$ be the frequencies of the different states present at the initial site, $Y_{02}, \ldots, Y_{k2}$ the frequencies at the next site, etc.
Different displacement policies can be applied in these kinds of experiment. One policy is to first move the needle until a site of activity is found, then displace it the same number of times in every experiment. Another policy is to successively move the needle from the first site of activity to a new site until there is no activity recorded. This last alternative yields recordings from a varying number of sites in different experiments.

Statistical Testing of no Modality Clustering

The hypothesis of no modality clustering is implied by assuming that all fiber states X_i are independent ran-

*) Prof. Ove Frank
Department of Statistics, University of Stockholm, Sweden

dom variables with the same distribution as X. Two test statistics that tend to take large values if there is type i modality clustering are

$$S_i = \sum_{j=1}^{r} Y_{ij} \text{ and } T_i = \sum_{j=1}^{r-1} Y_{ij}\, Y_{i,j+1} \text{ for } i = 1, \ldots, k$$

where r is the number of sites visited after and including the initial site of activity. Modality of any type tends to make the following statistics large:

$$S = \sum_{i=1}^{k} S_i, \quad T = \sum_{i=1}^{k} T_i, \quad U = \sum_{j=1}^{r-1} Z_j\, Z_{j+1}$$

where $Z_j = \sum_{i=1}^{k} Y_{ij} = m - Y_{oj}$. The probability distributions of these i=1 statistics under the hypothesis of no modality clustering can be determined empirically by simulating a large number of sequences $X_1, \ldots, X_{m+rh}$ and calculating the relative frequencies of different outcomes of the statistics involved.

An extension of the model which can be handled by a similar simulation method is obtained by allowing the displacements to be independent random variables with a common distribution. A simple displacement distribution of interest is uniform on the integers between 1 and h for some h between 1 and m.

References

[1] Hallin, R. (1990): Microneurography in relation to intraneural topography; Somatotopic organizat. of median nerve fascicles in man.
J. Neurol. Neurosurg. Psychiat. (to appear).

[2] Hallin, R. , Ekedahl, R. and Frank, O. (1990) : Segregation by modality of myelinated and unmyelinated fibers in human sensory nerve fascicles. **Muscle and Nerve** (to appear).

[3] Hallin, R. , Wiesenfeld-Hallin, Z. and Duranti, R. (1986): Percutaneous microneurography in man does not cause presure block of almost all axons in the impaled fascicles. **Neuroscience Letters** 68, 356-361.

AN ARCHITECTURE OF NEUROCOMPUTER FOR IMAGE RECOGNITION*)

*A. V. Gavrilov***)*

Abstract: A new kind of neurocomputer architecture for situation recognition and other complex image processing is proposed. The features of its neural elements and structure of the network is described. Some simulation results of this neural network are discussed.

In the last ten years, after the end of a great attention of the investigation in the neurocybernetics, initiated by Minsky M. and Papert S. [1], the attention to neural networks and neurocomputers was increased again. Particularly, this was stimulated by the design of Boltzman machine [2]. This attention is caused by the small ability of logistic approach for learning to recognize fuzzy things at first, and by appearing of new ideas which are able to give the direct recommendations for choice of the neural network structure and of their acting algorithms at second.

The applications of neural networks for image recognition are known for example, in the speech recognition [3] and in the control of industrial systems [4].

Almost all the known neural networks are constructed from very simple threshold elements, similar on formal neurons of McCulloch and Pitts. This paper deals with another approach to their construction based on the view of the elements of neural network as if the simple perceptrons learning to recognize the simple image — binary input code.

On the basis of this idea, the following architecture of neural network is proposed:

The elements (nodes) of network are connected by random links. Each its element ($i = 1, N$) has some binary inputs $u_{ij} \mid j = 1, M$ connected with the outputs of another elements, one binary input for adaptation a_i (A-input) and one output v_i. The state of the element is characterized by threshold h_i and key k_i. The rule of switching of its element is :

$$v_t = \begin{cases} 1, \text{ if } \sum_{j=1}^{M} f(u_{ij}, k_{ij}) > h_t \\ \\ 0 \text{ in other case,} \end{cases}$$

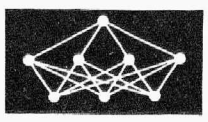
where k_{ij} — j th bit of key k_i ($j = 1, M$);
f — function equal *1* if $u_{ij} = k_{ij}$ and 0 in other case.

The vector of inputs $u_i = (u_{ij} ; j = 1, M)$ or the key may be represented by point in the code space. So the neural element may be represented as a perceptron recognizing the set of binary codes in any occurrence of key. The size of occurrence is determined by threshold. Therefore in this model the element of network executes more complex function hat in known neural network.

Some elements of neural network are connected to inputs by binary sensors of any kind. They compose the input layer. Others compose the output and hidden layers. The A-inputs are connected to outputs of another elements or to special sensors detecting the bad image or the other unsatisfactory situation in which the system, including the neural network, may be appeared. The goal of the neural network is to learn to escape from such bad situation.

*) Presented at the Latvian Signal Processing International Conference Proceedings. Riga, April 24-26, 1990, Vol. 2, p. 306-308.

**) Dr. A. V. Gavrilov, Department of Computers
Novosibirsk Electrotechnical Institute
K. Marx Str. 20, 630092 Novosibirsk 92, USSR



The neural network operates in the discrete time, i. e. for time the outputs of all elements determined by the inputs in the time $t = 1$.

For the design of model the behavior of the system coming-away from any moving object was chosen. The environment simulated by 8×8 points from the field where the system and object may move. For one step of time scale the object and the system can make one step in the space in any direction: up, down, to right or to left. Output layer of neural network includes four motor neurons. Signals on outputs of each of them indicate the move of the system in certain direction. The object and the system can not be situated in one point of field. The situation when the system is near the object is detected as bad. To delete a boundary effect the field is circled, i. e. the permanent linear moving is allowed.

Two kinds of the object behavior were simulated — linear moving (L) and random moving (R). In first case the moving in your direction is possible. The direction is changed, if the collision was not during the interval TT. For to compare the results of simulation, the probability of nearness of the system and the object P, computed during simulation time T was used.

The following parameters of neural network were changed:

— number of the nodes of network (N)
— number of the nodes in input layer (DR)
— per cent of A-inputs of the nodes connected with outputs of other nodes (PA)

Two kinds of connections between the input layer and the field were simulated. In first the elements of the input layer are connected with the points of the field and the signals on these inputs are determined by appearing of the object in corresponding points (G— global view). In second case four sensors are simulated. These sensors detect the relative position of the object: up, down, right or left from system. The inputs of the nodes of the input layer are connected with them. This approach may be called local view (L).

The program model is based on the following algorythm:

```
Input of parameters of model.
Creation of the neural network structure and its
  connection with sensors.
Set begin states of neural network and variables
  of model
WHILE t < T DO
  Show the picture of model state and held.
  IF t = TT THEN
    Computation of P.
    Reset of some variables of model
  ENDIF
  FOR i:=1 TO N DO
    Simulation of acting of i-th neural element.
  ENDFOR
  Simulation of step moving of system and
    object.
  t:=t + 1
ENDWHILE
```

Further some results obtained in experiments with program model for T=1000, TT=100 are shown.

Model	N	NR	PA	P	Number of network adaptation
LG	100	20	10	0.051	3
LG	100	48	10	0.044	4
LG	100	48	90	0.124	1
LG	100	30	10	0.097	2
LG	70	20	10	0.08	3
LG	50	20	0	0.044	4
LG	50	20	10	0.025	4
LG	50	30	90	0.14	0
LG	50	40	90	0.089	1
LL	100	48	10	0.032	4
LL	50	20	10	0.043	5
RG	100	40	10	0.096	
RG	50	20	10	0.46	
RG	50	30	10	0.08	
RG	50	20	50	0.082	
RG	50	20	70	0.071	
RG	50	20	90	0.08	

The results obtained by simulation allow to make the following conclusions:

— the proposed architecture provides the adaptation of acting of neural network in the changing and undetermined environment;
— in the case of linear moving of the object the best results were obtained for small per cent of connections of the A-inputs with outputs of another elements of network;
— this per cent not influence on effective of adaptation in the case of random behavior of object.

References

[1] Minsky M. , Papert S.: Perceptrons. MIT Press, Cambridge, MA, 1969.

[2] Ackley D. H. , Hinton G. E. , Sejnovski T. J.: A Learning Algorithm for Boltzmann Machines. Cognitive Science, 1985, Vol. 9, 147-169.

[3] Prager R. W. , Harrison T. D. , Fallside F.: Boltzmann Machines for Speech Recognition. Computer Speech and Language. 1986, No. 1, 3-27.

[4] Odri S. , Petrovacki D. , Grbovic J.: Neural Networks as a Part of Automatic Control System. In: Systems Science X:, The International Conference on Systems Science (Abstracts of papers), Wroclaw, 1989, 143-144.

A VIEW ON NEURAL NETWORKS PARADIGM DEVELOPMENT

J. Hořejš)*

The aim of this paper is to give the reader an introductory survey of the most known paradigms and technique, which he/she may encounter on his/her way either to theoretical research or practical applications of neural networks (NN[s]). We will not compete with textbooks and monographs with detailed and quite precise descriptions and elaborations. Instead we will rely on metaphors and simple examples, hoping that the reader at the assumed level (near to a beginner with basic knowledge of math) needs to catch the ideas and flavor rather than being overwhelmed by intricacies. In the presentation, we shall loosely follow the history, starting briefly with the "first generation" of NNs at the end of the Second Word War, introducing then "second generation" perceptrons and ending with the promises of the current "third generation". Yet we tried to introduce basic concepts as quickly as the text methodically permits. The generations are characterized not only by successive enrichments of new concepts and techniques; they are also separated in time by about twenty years gaps, demonstrating successes and failures and at the same time successive shifts from biological considerations (brain research, if you like) to non-traditional information processing machines ready to serve as new tools of the computer science community. Although we often prefer neurophysiological terminology (thus speaking about "neurons" instead of more indifferent "processing elements" etc.) which suggests a lot of useful psychological and similar metaphors, the final aim is to present a technically oriented account.

There are now thousands of contributions on NNs and maybe hundreds of them should be indeed read depending on reader's interest and ability. What to recommend in such a situation to a beginner? Try to follow this tutorial; if you consider it legible and the topic catches you, proceed with some elementary textbook [Wasserman's Neural Computing — Theory and Practice, van Reinhold Nostrand 1989 is a hint] and then by some of the first monographs (like Hecht-Nielsen's Neurocomputing, Addison Wesley 1989). During this introductory exposition we shall not bore you by suggestions to get and read anything else; if you like, just remember the names scattered through out the text — they are mostly famous by now. And at the end of the tutorial we bring a selected, briefly annotated bibliography.

A NN is generally an oriented graph, the nodes of which are so called *neurons* [processing elements, cells] and arcs are *connections* between them. Using neurophysiological analogy we often speak about axons: axon is the only connection leaving a neuron body [soma] possibly branching to many *terminals*; on the end of each there resides a so called synapse, which mediates the *signal* [impulse] spreading over the axon further on. Each neuron of a net gathers some *impulses [stimuli]* from other neurons or from the environment and under certain circumstances (when it is stimulated enough) it responds by another stimulus, sending it along the axon, and thus the terminals and synapses, to the rest of the net (i. e. to some other neurons) or to the external word according to proper *activation dynamics laws*. If a neuron is in this way *activated, it "fires"*, and its activity is passed over its axon. If the axon branches, all terminal synapses receive the same stimulating power — the energy of the stimulus does not decrease. By this mechanism a *spreading activity* in the whole net is provoked.

Under certain conditions, the whole net can change, mostly by changing the *strengths* of individual connections [*weights* of terminal synapses] coming out of a neuron to other neurons [specifically to their morphological parts called dendrites]. The efficiency (weights) of synapses can be characterized numerically and under certain conditions these numbers can be modified — the net *adapts* itself in order to perform a given task more suitably. The way how such an adaptation is performed is given by some *adaptation dynamics laws*. In living organisms (and some models), both activation and adaptation processes run concurrently. In many simpler models and in computer simulation we distinguish between *active [working] mode*, in which the net (already) satisfactorily does its job for which it has been designed, and *adaptive [selforganizing] mode*, in which it is tuned for its own improvements. We shall not meet adaptation until section 3, where its importance and characteristics will become a supreme topic.

A real brain contains about 10^{11} neurons and about 10^{14} synapses. *Neurocomputers*, which are artificial devices for simulating NNs (ranging from usual PC's to sophisticated experimental devices), deal at present with hundreds to billions of neurons and synapses.

1. NN as a calculus for nervous activity [The first generation].

The first (formal) neurons of McCulloch and Pitts were very simple indeed. Two sorts of axons were permitted to connect to them: *excitatory* (depicted by full dots in figures) and *inhibitory* (empty dots). We speak also about excitatory or inhibitory synapses. Every neuron has its *threshold* (indicated by the number in-

*) Prof. Dr. Jiří Hořejš, Department of Computer Science, Charles University, 11800 Prague 1, Malostranské nám. 25, Czechoslovakia.

scribed inside the node). Whenever excitations are greater (in their number, say) than inhibitions so that the difference exceeds the neuronal threshold, the neuron becomes activated and sends its own impulse over its axon to the rest of the net. *Fig. 1* demonstrates how easy it was to model trivial conditioning. If an external unconditional stimulus U came to the net, the leftmost bottom neuron, and afterwards also the output (top) neuron, fired and caused some active response R of the net. If, at some time moment (for simplicity we consider discrete time scale), both U and the conditional stimulus C appear concurrently, both intermediate neurons are activated (note the values of thresholds!) and one time step later the top neuron fires as well. Now, if immediately C alone appears, then due to the stimulus C and the feedback loop of the rightmost bottom neuron (representing memory of previous activity), this neuron fires too and so does the top one — the same effect occurs.

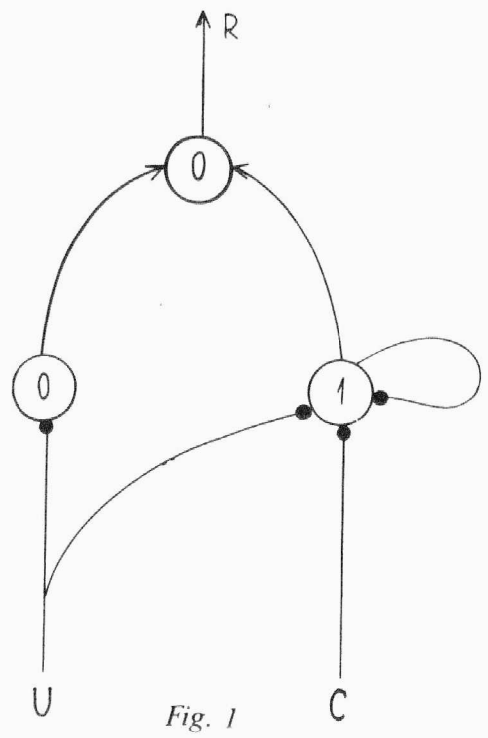


Fig. 1

The net of *Fig. 2* is slightly more complicated. Unlike the first case, now there is also an inhibitory synapse and the condition for activation of every neuron now reads

$$\Sigma z_i^+ - \Sigma z_i^- > \vartheta; \qquad (1)$$

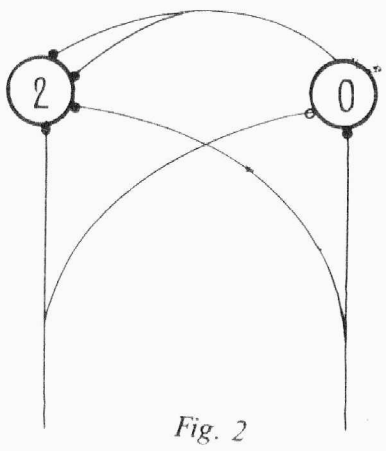


Fig. 2

here ϑ is the threshold of the considered neuron, $\mathbf{z}$ is a vector consisting of impulses on all connection axons (terminals) leading to the neuron, $\mathbf{z} = [z_1, z_2, \ldots, z_m]$, and every connection z_i is either activated ($z_i = 1$) or not ($z_i = 0$). The formula (1) is our first example of a simple activation dynamics law.

For a z_i^+ to contribute to the first sum in (1) it is inecessary and sufficient that $z_i = 1$ and the i-th axon have an iexcitatory synapse. Similarly for the second sum. Now let $\mathbf{x} = [x_1, x_2]$ be an input vector to the net and $\mathbf{a} = [a_1, a_2]$ the vector of current activities of the net, a_i representing the activity (1 or 0) of the neuron i. Under influence of $\mathbf{x}$, vector $\boldsymbol{a}$ generally changes. Because there are four possible input vectors, four possible activity (also „state") vectors and because — above all — the state vector (initially e. g. $\boldsymbol{a} = [0, 0]$) changes dynamically in time, it is not so easy to predict immediately the net behavior under a changing environment, represented by a sequence of external input vectors. To see what may happen, we help ourselves by drawing the transition graph of the net, establishing beforehand what happens if in a given state we encounter any of the possible inputs. This graph is given in *Fig. 3*. It consists of four boxes; within a box the state $a_1 a_2$ is inscribed. Any of the 16 arrows is labeled by one or more pairs $x_1 x_2$. Given a state (box) and an input vector, we find out the arrow labeled by that input vector and receive at its end the next state. The reader surely recognizes that this is a *finite automaton;* as a matter of fact the idea of finite automata, so useful in computer science, originated by studying NNs.

Now we generalize the model of a neuron a bit.

2. Perceptrons [Second generation]

First of all we extend the binary data used till now by admitting that both incoming stimuli and synapses may acquire the arbitrary real values. With z_i expressing the strength of i-th stimulus and w_i the weight of the synapse, we change the activation dynamics (1) to

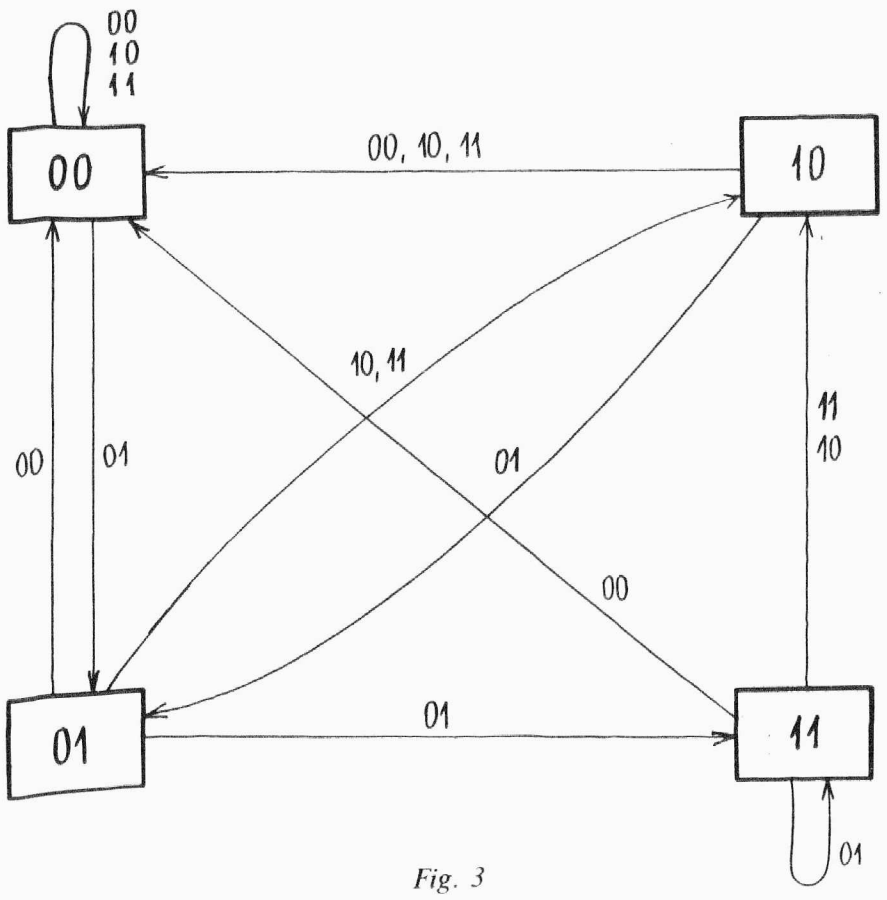


Fig. 3

$$y = \sum_{i=1}^{m} w_i z_i - \vartheta \qquad (2)$$

If $w_i > 0$ we speak again about excitatory synapses, if $w_i < 0$ the term inhibitory synapse fits. The neuron now fires if and only if the total sum of stimuli, each recalculated w. r. t. its synaptical weight, exceeds the threshold. The output of the neuron is now also real-valued. Because we shall use it first as a classifier, which accepts some input vectors **x** and rejects others, we introduce *A* convention that actual output is again binary, assuming a value 1 (representing the answer YES to the classification problem) if the right hand side of (2) is greater than 0 (the neuron fires), and 0 (representing NO) otherwise. This can be achieved by introducing another, nonlinear function *S (signum)*, defined simply by $S(\eta) = 1$ whenever $\eta > 0$ and $S(\eta) = 0$ otherwise. By introducing a fictive neuron with constant output -1 and setting $w_0 = \vartheta$, we can rewrite (2) in a more homogeneous form as

$$y = S\left(\sum_{i=0}^{m} w_i z_i \right) \qquad (3)$$

which can be depicted as in *Fig. 4*, often omitting

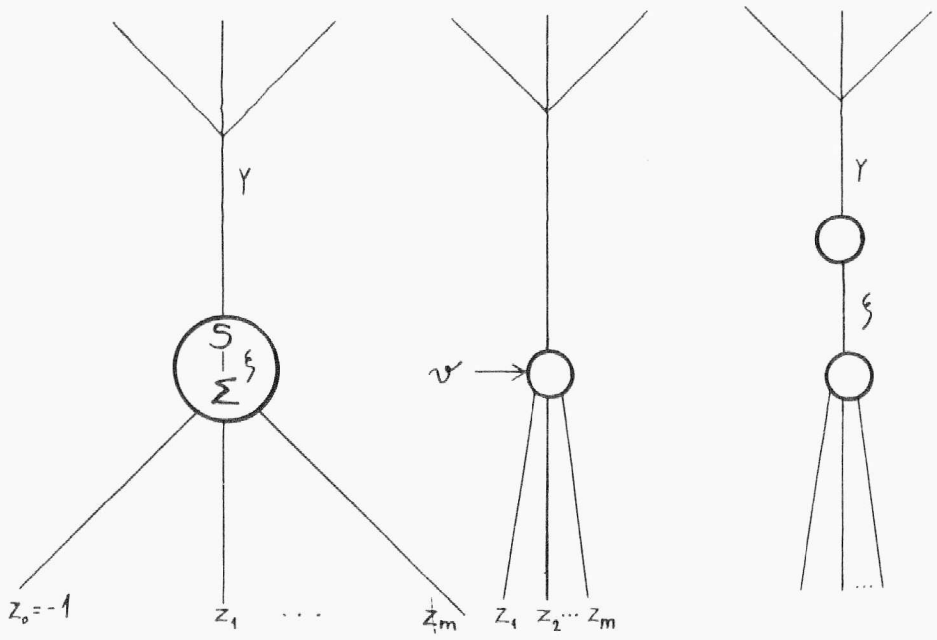


Fig. 4

Σ and *S* inside the circle or splitting the neuron into two twins and thus the linear and nonlinear part of the mapping (3). The sum $\zeta = \Sigma\, w_i x_i$ is occasionally called the *net income* of the neuron.

Note carefully that $\Sigma\, w_i z_i$ is a linear form, $\Sigma\, w_i z_i = 0$ is an equation of a hyperplane (for $m = 2$ a straight line in a usual two dimensional plane). w_0 under the last convention represents then the offset (shift from the origin of cartesian coordinates), remaining w_i's form the normal vector of that hyperplane. The neuron fires if and only if the weighted sum of all nonfictive contributors exceeds the threshold $\vartheta = w_0$. For a given vector **w** (which we standardly interpret as the vector specifying a hyperplane) (3) separates all vectors **z** into two categories: if $y = 1$, the vector z lies in the positive m-dimensional halfspace 2 and specifies

and specifies a point which is at the distance $\Sigma\, w_{i\,zi}$ / sqrt $(\Sigma\, w_i^2)$, where the first sum is taken for $i = 0, 1, \ldots, m$ and the second one only for $i = 1, \ldots, m$. If the scope of indices is clear from the context, we often abbreviate the sum $\Sigma\, w_i x_i$ by the dot product **w . x**.

If we now characterize all objects from an *input space (stimuli environment)* by real vectors (coordinates of which numerically describe various symptoms/tags), the described form of neuron — *perceptron* can separate and thus *recognize* two *categories:* those which are in the positive halfspace giving $y = 1$ and those which are not ($y = 0$). If the symptom vectors are treated like descriptions of the first category, the perceptron will fire, in the remaining cases it remains still. Actually perceptrons were introduced just for recognition of perceptional stimuli (visual, audio) corresponding to some interesting objects and distinguishing them from the others. It should be noted that the vector **x** can describe a very concrete form of a visual information [e. g. bit valued (0/1 = white/black) pixels of a matrix raster read line after line similarly to a TV screen or a retina], but it may be composed of any quantitatively expressed abstract data characterizing external object[s] [e. g. in medicine we can have **x** = [BP (blood pressure), FW (sediment), No of leukocytes, . . .].

Similarly a perceptron assigned a task of recognizing correct signatures from forgeries (GOOD objects from BAD ones) written on a raster $p \times q$ might use as an input the binary vector $\mathbf{x} = [x_1, \ldots, x_{pq}]$, where $x_{(p-1)\,i+j}$ means that the square in the i-th row and j-th column intersects the signature line; or it may (better) use a real input vector **x**, the components of which describe such aspects as duration of the process of signing, maximal acceleration achieved during it, number of lifts of the pen, etc.

Imagine now a highly metaphoric "brain" of a chicken, consisting of a single perceptron only. Receptory organs (eyes, ears) supply the chicken by a many-dimensional stimuli vector **x** (including some characteristics of shape and noise of an approaching object, etc). Let the chicken have its perceptron (specified by the vector **w**) originally established by innate genetic information so that if the approaching object x is interpreted as a hawk (squares in *Fig. 7a*), the perceptron fires and starts a sequence of escape actions, while if **x** "reminds" it of a farmer (examples of which are depicted by circles), the neuron and the chicken remain still. So the perceptron again classifies the input vector space into categories BAD and GOOD. For an $x \in$ BAD, it is **w. x** > 0, while for an **x** $\in$ GOOD, **w . x** ≤ 0.

As the reader hopefully noticed, one perceptron is only able to distinguish between two subsets of the input space, which are separable by some hyperplane — they have to be *linearly separable*. The simplest case of an impossibility of solving (by a single neuron!) a linearly nonseparable problem is the implementation exclusive OR, called XOR, where [00], [11] $\in$ GOOD, while [01], [10] $\notin$ GOOD. [Try to draw a straight line

separating the two pairs of vertices in the square!]. *The fact is however that by introducing more neurons this limitation can be overcome, as will be shown immediately.*

So let us see how to avoid the necessity of linearly separable subsets; at the same time it should be clear that we do not need to restrict ourselves to a binary classification (YES/NO, GOOD/BAD). We shall first introduce the concept of a *multilayered* NN. Let there be several layers (subsets) of neurons (at present perceptron-like, i. e. governed by equation (3)). The layers are imagined to be ordered in a vertical fashion. The bottom layer serves for an input of say m-dimensional *input vectors*. The upmost *n*-dimensional layer gives the *output vectors*. In between these two there may be several so-called *hidden layers*. Neurons of a layer connect (only) to neurons in the layer above (look at *Fig. 6* for an example). Please notice now that with these sorts of multilayered nets we are able to implement boolean functions: given one or more neurons, we can connect their outputs to another neuron (in a layer above, with synaptical weights 1) so that the above neuron fires if and only if α) all lower neurons fire (generalization of boolean AND); β) at least one lower neuron fires (generalization of OR); γ) none of the lower neurons fires (generalization of NOT). To see α), simply choose as the threshold of the above neuron one less than is the number of lower neurons, to see β), choose the threshold 0, to see γ), choose the threshold -1.

Now let a subset GOOD of the input space be convex, i. e. there are several hyperplanes such that every $\mathbf{x} \in$ GOOD lies in their positive halfspaces. Assign to every hyperplane a neuron (in *Fig. 5a* we have a triangle, members of GOOD are squares, separating lines and corresponding neurons connected by dotted lines) and create above them another neuron which fires if and only if all of the externally stimulated neurons do, which happens exactly if input $\mathbf{x}$ belongs to GOOD. So we can characterize by the topmost neuron the membership to any convex subset. In a layer higher still we can similarly create neuron characterizing unions (actually any boolean function) of such convex sets — see *Fig. 5b*. And finally, because we have *n* neurons in the topmost layer, we can categorize the input space into *n* different subsets. In conclusion, three layers are enough for such a categorization.

Following this general algorithm we can however arrive at a NN which may be rather cumbersome if the members of, say, GOOD objects are scattered through the input space rather irregularly, requiring in the worst case the treatment of any member of GOOD by a special "small triangle". That this need not always be the case is shown in the last example of this section. Now we describe how much more elegant solutions can often be achieved, and later on we will see that such solutions can be reached even by an automatic adaptation of the net.

Let the input space consist of all binary 6-dimensional vectors and include in GOOD exactly those which are center — symmetric. Thus e. g. $\mathbf{x} = [110011] \in$ GOOD, while $\mathbf{x} = [101100] \notin$ GOOD. The net of *Fig. 6* solves this problem with only 6+2+1 neurons, where, moreover, the input layer performs no computation, just fans-out the information from the input vector to the neurons in the layer above. Note that the weights from the input layer to the hidden one are powers of 2, so that any combination of the (first) three of them cannot compensate the total "income" of the remaining three in the case that the input vector is not symmetric. [If $\mathbf{x}$ is symmetric, the net income of both hidden neurons is 0 and both fire; otherwise there are two symmetrically positioned neurons, one of which — say x_2 — has the value 1 and the other — here x_5 — has the value 0. In this case the right hidden neuron lacks the compensating contribution 2 from x_5 and receives thus the net income which does not exceed its threshold.] This gives a hint of how to construct the solution of an *symmetry problem* for an arbitrary *m*-dimensional case (*m* even).

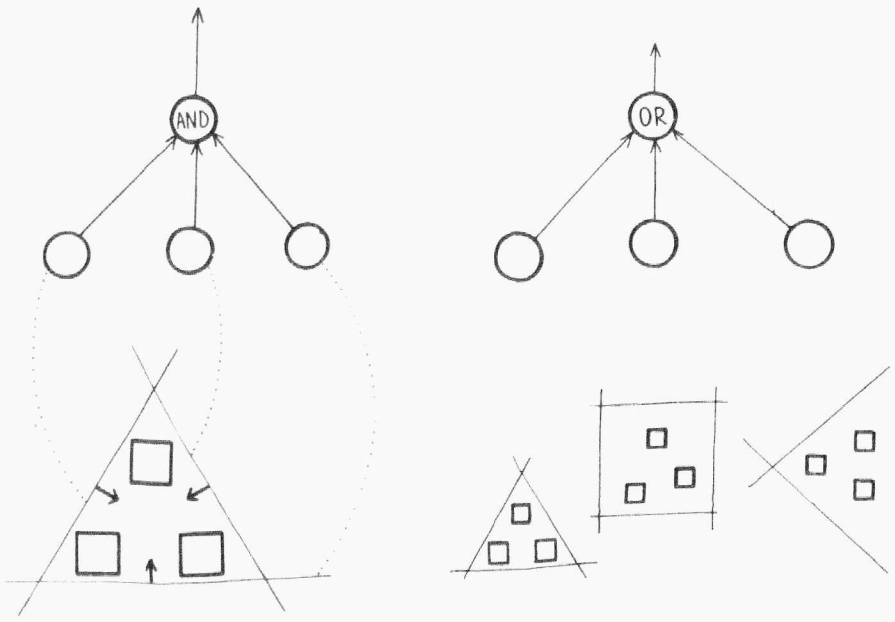


Fig. 5

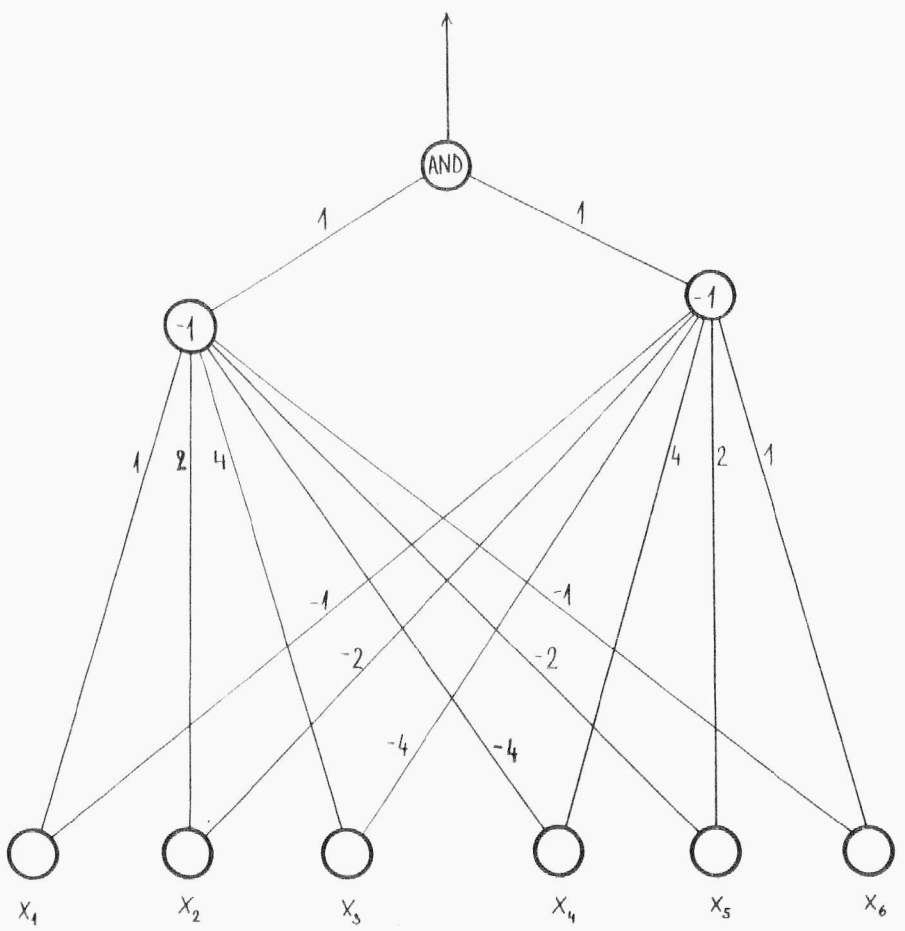


Fig. 6

Instructions to authors

1. Manuscript
Two copies of the manuscript should be submitted to the Editor-in-Chief.

2. Copyright
Original papers (not published or not simultaneously submitted to another journal) will be reviewed. Copyright for published papers will be vested in the publisher.

3. Language
Manuscripts must be submitted in English

4. Text
Text (articles, notes, questions or replies) double space on one side of the sheet only, with a margin of at least 5 cm, (2") on the left. Any sheet must contain part or all of one article only. Good office duplication copies are acceptable. Titles of chapters and paragraphs should appear clearly distinguished from the text.
Author produced (camera ready) copy is acceptable if typed on special sheets which are available from the Editor, and adherence to the Typing instructions (also available from the Editor) has been taken care of, is emphasized that camera ready text should be typed *single* space (i. e. with *no* space between the lines).
Complete text records on 5 1/4" floppy discs are prefered, if typed according to the instructions available from the Editor.

5. Equations
Mathematical equations inserted in the text must be clearly formulated in such a manner that there can be no possible doubt about meaning of the symbols employed.

6. Figures
The figures, if any, must be put on separate sheets, clearly numbered and their position in the text marked. They must be drawn in Indian ink on white paper or tracing paper, bearing in mind that they will be reduced to a width of either 7,5 or 15 cm (3 or 6") for printing. After scaling down, the normal lines ought to have a minimum thickness of 0,1 mm and maximum of 0,3 mm while lines for which emphasis is wanted can reach a maximum thickness of 0,5 mm. Labelling of the figures must be easy legible after reduction. It will be as far as possible placed across the width of the diagram from left to right. The height of the characters after scaling down must not be less than 1 mm. Photographs for insertion in the text will be well defined and printed on glossy white paper, and will be scaled down for printing to a width of 7,5 to 15 cm (3 to 6"). All markings on photographs are covered by the same recommendations as for figures. It is recommended that authors of communications accompany each figure or photograph with a descriptive title giving sufficient information on the content of the picture.

7. Tables
Tables of characteristics or values inserted in the text or appended to the article must be prepared in a clear manner, preferably as Camera Ready text. Should a table need several pages these must be kept together by sticking or other appropriate means in such a way as to emphasize the unity of the table.

8. Summaries
A summary of 10 to 20 typed lines written by the author in the English will precede and introduce each article.

9. Required information
Provide title, authors, affiliation, data of dispatch and a 100 to 250 word abstract on a separate sheet. Provide a separate sheet with exact mailing address for correspondence

10. Reference
References are recommended to be listed alphabetically by the surname of the first author. List author(s) (with surname first), title, journal name, volume, year, pages for journal references, and author(s), title, city, publisher, and year for the book references. Examples for article and book respectively:

Dawes, R. M. and Corrigan, B.: Linear models in decision making, Psychological Bulletin, **81** (1974), 95—106

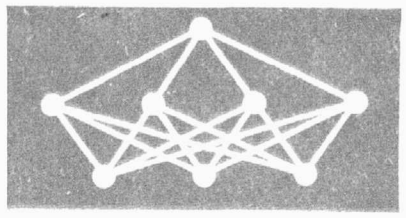
Brown, R. G.: Statistical Forecasting for Inventory Control, New York: McGraw-Hill, 1959.

All references should be indicated in the manuscript by the respective number or by the author's surname followed by the year of publication (e. g., Brown, 1959).

11. Reprints
Each author will receive 25 free reprints of his article.

INTERNATIONAL SCIENTIFIC CONFERENCE

IDG



"REALITY AND TRENDS IN NEUROCOMPUTING"

sponsored by the IDG Co., Czechoslovakia
Brno, October, 23.–24

The conference will have the following three scientific sections:
1. Artificial Neural Networks — Theory and Applications
2. Neurocomputers — A New Tool for Informatics
3. Trends in Neurocomputing

Among the invited speakers are expected:
prof. Kerckhoffs, Netherlands
prof. Haken, Germany
prof. Gupta, Canada
prof. Frolov, USSR
prof. Taylor, GB
prof. Koruga, Yugoslavia
prof. Hornik, Austria
prof. Dreyfus, France
doc. Faber, Czechoslovakia
doc. Hořejš, Czechoslovakia
prof. Dudziak, USA
prof. Cimagalli, Italy
prof. Marko, Germany

Contact adress:
dr. Mirko NOVÁK — chairman of the conference
Institute of Computer and Information Science
Pod vodárenskou věží 2
182 07 PRAGUE 8
Czechoslovakia

Phone: (00422) 815 2080, (00422) 82 1639
Fax: (00422) 858 5789
E-mail: CVS 35 @ CSPGCS 11. BITNET

The scope of the conference is to be a free forum for presentation of new ideas and exchange of meanigs on the:
— new views on the theoretical and practical problems of neuroscience with special regard to neurocomputing,
— role of neurocomputers in the process of information processing,
— present and new application of neurocomputing in science, engineering, industry and commerce,
— expected development of neurocomputers.

Language: English

The program of each section will involve the blocks of talks presented by invited speakers with enough space for discussions. The conference will be closed by the overall panel discussion.

Conference fee: 499 US$

Payments: by bank transfer to Czechoslovak Business Bank, Praha 1
account No. 3483—39129 (for US$)
account No. 3427—39129 (for DEM)
by check payable to Computer World Co.
Blanická 16
120 00 PRAGUE 2
Czechoslovakia

Accomodation: in Brno hotels and privats

hotel *****	(210 US$ per night)
hotel ****	(180 US$ per night)
hotel ***	(80 US$ per night)
privat	(35 US$ per night)